

INVITED PAPER

For the Special Issue: The Evolutionary Importance of Polyploidy

# Strongly asymmetric hybridization barriers shape the origin of a new polyploid species and its hybrid ancestor<sup>1</sup>

Mario Vallejo-Marín<sup>2</sup>, Arielle M. Cooley<sup>3</sup>, Michelle Yuequi Lee<sup>2</sup>, Madison Folmer<sup>5</sup>, Michael R. McKain<sup>4</sup>, and Joshua R. Puzey<sup>5,6</sup>

**PREMISE OF THE STUDY:** Hybridization between diploids and tetraploids can lead to new allopolyploid species, often via a triploid intermediate. Viable triploids are often produced asymmetrically, with greater success observed for “maternal-excess” crosses where the mother has a higher ploidy than the father. Here we investigated the evolutionary origins of *Mimulus peregrinus*, an allohexaploid recently derived from the triploid *M. xrobertsii*, to determine whether reproductive asymmetry has shaped the formation of this new species.

**METHODS:** We used reciprocal crosses between the diploid (*M. guttatus*) and tetraploid (*M. luteus*) progenitors to determine the viability of triploid *M. xrobertsii* hybrids resulting from paternal- vs. maternal-excess crosses. To investigate whether experimental results predict patterns seen in the field, we performed parentage analyses comparing natural populations of *M. peregrinus* to its diploid, tetraploid, and triploid progenitors. Organellar sequences obtained from pre-existing genomic data, supplemented with additional genotyping was used to establish the maternal ancestry of multiple *M. peregrinus* and *M. xrobertsii* populations.

**KEY RESULTS:** We found strong evidence for asymmetric origins of *M. peregrinus*, but opposite to the common pattern, with paternal-excess crosses significantly more successful than maternal-excess crosses. These results successfully predicted hybrid formation in nature: 111 of 114 *M. xrobertsii* individuals, and 27 of 27 *M. peregrinus*, had an *M. guttatus* maternal haplotype.

**CONCLUSION:** This study, which includes the first *Mimulus* chloroplast genome assembly, demonstrates the utility of parentage analysis through genome skimming. We highlight the benefits of complementing genomic analyses with experimental approaches to understand asymmetry in allopolyploid speciation.

**KEY WORDS** allopolyploidy; asexual reproduction; chloroplast genome; genome skimming; hybridization; introduced species; *Mimulus peregrinus*; mitochondrial genome; sterile hybrid; triploid block

The extraordinary abundance of polyploid lineages among flowering plants (Otto and Whitton, 2000; Soltis et al., 2014; Barker et al., 2016) has led to great interest in understanding the processes of polyploid formation and establishment in the wild. Polyploids (individuals carrying more than two sets of chromosomes) can originate through a variety of pathways. These include the somatic doubling of chromosomes in meristematic tissue during mitosis

and, perhaps more commonly, unreduced gametes (Ramsey and Schemske, 1998; Mallet, 2007). When the formation of polyploids involves fusion of gametes with different ploidy levels (e.g., reduced and unreduced gametes, or between a diploid and a tetraploid), an individual with an unbalanced ploidy level (e.g., a triploid) will be formed (Levin, 2002). In most cases, these odd-ploidy individuals are a transient phase in polyploid formation, leading to the eventual production of higher ploidy, stable lineages such as tetraploids and hexaploids (Levin, 2002).

Odd-ploidy levels represent a significant challenge to overcome in the formation of polyploids because interploidy crosses often have strong postzygotic reproductive barriers, severely reducing hybrid viability and fertility (Ramsey and Schemske, 1998; Levin, 2002; Comai, 2005; Köhler et al., 2010). The term “triploid block” has been coined to describe the regular observation that interploidy

<sup>1</sup> Manuscript received 6 November 2015; revision accepted 16 February 2016.

<sup>2</sup> Biological and Environmental Science, School of Natural Sciences, University of Stirling, Stirling, Scotland, FK9 4LA UK;

<sup>3</sup> Biology Department, Whitman College, Walla Walla, Washington 99362 USA;

<sup>4</sup> Donald Danforth Plant Science Center, St. Louis, Missouri 63132 USA; and

<sup>5</sup> Department of Biology, College of William and Mary, Williamsburg, Virginia 23185 USA

<sup>6</sup> Author for correspondence (e-mail: [jrpuzey@wm.edu](mailto:jrpuzey@wm.edu))

doi:10.3732/ajb.1500471

crosses, specifically between diploids and tetraploids, often result in early hybrid inviability and sterility (Köhler et al., 2010). However, evidence from both natural and experimental systems has shown that interploidy crosses can sometimes produce odd-ploidy offspring that are viable (Burton and Husband, 2000; Husband, 2004; Stace et al., 2015).

The ability to overcome the triploid block can be asymmetric, with offspring viability dependent on whether the maternal or paternal parent had the higher ploidy level (Köhler et al., 2010). Asymmetric hybridization has considerable potential for influencing population genetic processes. It generates a predictable bias in cytoplasmic inheritance and, in some cases, can promote unidirectional introgression (Field et al., 2011). Repeated backcrossing between an asymmetrically produced hybrid and its paternal pollen source may lead to “cytoplasmic capture,” in which the mitochondria and chloroplasts of the maternal progenitor are combined with the nuclear genome of the paternal progenitor (Petit et al., 2004). Asymmetric introgression can have conservation implications, for example, through genetic swamping or competitive displacement of the rarer progenitor species (Rhymer and Simberloff, 1996).

Asymmetric hybridization can be explained by a variety of mechanisms, including cytonuclear incompatibilities (Tiffin et al., 2001) or imbalances in either the ploidy ratio of endosperm to embryo (Watkins, 1932; Burton and Husband, 2000) or between maternal to paternal genomes (Haig and Westoby, 1989; Köhler et al., 2010). Barriers to interploidy mating within and between species, including triploid block, have been studied using controlled crosses in a variety of natural and artificially produced diploids and polyploids (Ramsey and Schemske, 1998; Husband, 2004; Köhler et al., 2010; Scott et al., 2013). Earlier surveys of interploidy crossing data determined that maternal-excess crosses—those in which the maternal parent has a higher ploidy level than the paternal one—are consistently more successful than paternal-excess crosses at producing viable triploid seeds (Stebbins, 1957; Ramsey and Schemske, 1998). However, a number of counter-examples, in which paternal-excess crosses are more successful, have been recently discovered (Table 1; Appendix S1).

Asymmetric reproductive barriers during interploidy hybridization can be identified experimentally by performing reciprocal crosses (e.g., Husband, 2004), but discovering whether asymmetries observed in the laboratory translate to patterns in nature requires tracing the maternal vs. paternal ancestry of wild populations. Ancestry analysis of this kind requires uniparentally inherited genetic markers, such as mitochondria or chloroplast that contain enough genetic differences between parental taxa. The rapid growth of next-generation sequencing projects offers an extraordinary opportunity to generate (sometimes inadvertently) exactly the type of data necessary to elucidate the maternal ancestry of wild populations at a genomic scale. Because cytoplasmic genomes (mitochondria and chloroplast) are present at a much higher number per cell than the nuclear genome, targeted and nontargeted sequencing often results in incidental high coverage of cytoplasmic genomes, as well as of repetitive elements in the nuclear genome (Dodsworth, 2015). Thus, mitochondrial and chloroplast information can be “skimmed” from the larger collection of sequences obtained in whole-genome sequencing projects (Straub et al., 2012) or from targeted sequencing approaches (Weitemier et al., 2014) and used for parentage analyses. These resource-efficient approaches not only rescue information that would otherwise be discarded, but also provide effective means for addressing longstanding

questions about the ancestries of naturally occurring polyploid lineages.

A new system for investigating allopolyploid speciation, using the complementary approaches of experimental manipulation and next-generation sequencing, is the complex of naturalized *Mimulus* L. (Phrymaceae) species in Great Britain and Ireland. Hybridization between introduced populations of the North American diploid *M. guttatus* ( $2n = 2x = 28$ ) and the South American tetraploid *M. luteus* ( $2n = 4x = 60\text{--}62$ ) have given rise to *M. ×robertsii*, a triploid ( $2n = 3x = 44\text{--}46$ ), sexually sterile, but widely naturalized taxon distributed across the British Isles (Roberts, 1964; Parker, 1975; Silverside, 1990; Vallejo-Marín et al., 2015). By means of rapid vegetative propagation, *M. ×robertsii* has become common and is now found in about 40% of British *Mimulus* populations (Vallejo-Marín and Lye, 2013). The reproductive asymmetry typical of interploidy matings has been reported for *M. guttatus* × *M. luteus* hybrids, but in the less common direction: triploids have been successfully produced only by paternal-excess crosses, with tetraploid *M. luteus* as the pollen donor (Roberts, 1964; Parker, 1975). These early studies suggest that *M. ×robertsii* is likely to have arisen in the same way, from a *M. guttatus* mother and *M. luteus* father. *Mimulus ×robertsii* has given rise at least twice to a sexually fertile allohexaploid species, *M. peregrinus* Vallejo-Marín ( $2n = 6x = 92$ ), through genome doubling (Vallejo-Marín, 2012; Vallejo-Marín et al., 2015). This speciation event is quite recent, as both progenitors were introduced to the United Kingdom (UK) in the last 200 years, and the earliest records of hybrid populations date back approximately 140 years (Preston et al., 2002; Vallejo-Marín et al., 2015). Given that *M. peregrinus* evolved from *M. ×robertsii*, one would expect that the maternal ancestor of *M. peregrinus* is also *M. guttatus*.

Here we investigated the occurrence and consequences of interploidy reproductive barriers in polyploid *Mimulus* (*M. ×robertsii* and *M. peregrinus*) produced by hybridization between diploid (*M. guttatus*) and tetraploid (*M. luteus*) taxa. We complemented classic crossing experiments with analyses of the parentage of natural hybrid populations using genome-skimming of chloroplast and mitochondrial genomes and single-marker genotyping. Interploidy hybrids in *Mimulus* are ideally suited for this purpose due to their experimental tractability (Wu et al., 2008) as well as the growing genomic resources available for this genus, including whole-genome sequences for both nuclear and mitochondrial genomes for *M. guttatus* (Mower et al., 2012; Hellsten et al., 2013), ongoing work on the *M. luteus* genome (P. P. Edger [Michigan State University] et al., unpublished manuscript), and recent whole-genome and targeted sequence data for native and introduced *Mimulus* populations (Puzey and Vallejo-Marín, 2014; Vallejo-Marín et al., 2015). We experimentally evaluated the success of reciprocal crosses between *M. guttatus* and *M. luteus* to confirm previous observations of triploid block asymmetry and compared the phenotype of hybrids produced in reciprocal directions. We used a combination of genome skimming and genotyping of natural populations to evaluate the hypothesis that reproductive asymmetry observed between *M. guttatus* and *M. luteus* violates the trend (Ramsey and Schemske, 1998) of triploid formation via maternal-excess crosses. Our study addressed the following specific questions: (1) What is the effect of cross direction on the phenotype and viability of hybrids? (2) Are naturalized triploid hybrids between *M. guttatus* and *M. luteus* (i.e., *M. ×robertsii*) asymmetrically produced, and does the asymmetry match the pattern observed in experimental hybridizations? (3) What is the maternal parent of

**TABLE 1.** Recent (1998–2015) reports of asymmetric crossing success following interploidy crosses. Some of the studies listed below verified that the seeds or seedlings produced were triploid; others did not and could potentially include diploid, tetraploid, or aneuploid offspring in their measure of success.

| Genus                            | Reference                     | Ploidy type                                | Superior direction | Measure of success   |
|----------------------------------|-------------------------------|--|--------------------|--|
| <i>Brassica</i>                  | Stoute et al., 2012           | auto (a)                                   | m                  | Seed germination (%); all seedlings tested were triploid   |
| <i>Brassica</i> × <i>Sinapis</i> | Cheung et al., 2015           | allo                                       | m                  | Triploid seedlings per pollination, relative to intraploidy crosses  |
| <i>Buxus</i>                     | Van Laere et al., 2015        | unknown                                    | m                  | Seedlings per pollination  |
| <i>Calluna</i>                   | Behrend et al., 2015          | auto (n)                                   | m                  | Triploid seedlings per pollination   |
| <i>Erythronium</i>               | Roccaforte et al., 2015       | auto (n)                                   | m                  | Seeds per pollination relative to intraploidy crosses  |
| <i>Lonicera</i>                  | Miyashita and Hoshino, 2015   | unknown                                    | m*                 | Triploid seedlings per pollination   |
| <i>Oryza</i> (rice)              | Sekine et al., 2013           | auto (u)                                   | m                  | Seed germination (%)   |
| <i>Acacia</i>                    | Nghiem et al., 2013           | auto (c)                                   | p                  | Filled fruits per pollination and filled seeds per fruit, relative to intraploidy crosses                        |
| <i>Aster</i>                     | Castro et al., 2011           | autohexaploid (n)                          | p                  | Seeds per pollination, seed germination (%), triploid seedlings per pollination, relative to intraploidy crosses |
| <i>Castilleja</i>                | Hersch-Green, 2012            | unknown naturally occurring 2x, 4x, and 8x | p                  | Filled fruits per pollination and seeds per fruit, relative to intraploidy crosses                               |
| <i>Chamerion</i>                 | Sabara et al., 2013           | unknown naturally occurring 2x and 4x      | p                  | Observed vs. expected triploid production from wild-collected maternal seed families                             |
| <i>Leucanthemum</i>              | Greiner and Oberprieler, 2012 | allo                                       | p                  | Seeds per pollination relative to intraploidy crosses  |
| <i>Solanum</i>                   | Dinu et al., 2005             | auto (u)                                   | p*                 | Seed development and seed germination (%)  |

Notes: a, derived from anther culture; c, colchicine-induced; n, naturally occurring; u, type of autotetraploid unknown; m, crosses more successful when maternal parent has higher ploidy; p, crosses more successful when paternal parent has higher ploidy. \*With embryo rescue.

hexaploid *M. peregrinus*, and does this fit the pattern seen for experimental crosses and naturalized *M. xrobertsii*?

## MATERIALS AND METHODS

**Study system**—The introduction of *Mimulus* into Europe in the 1800s resulted in hybridization between closely related, but previously geographically isolated taxa. Hybridization between taxa in the North American *M. guttatus* DC. and South American *M. luteus* L. species aggregates has yielded a taxonomically complex array of hybrids that establish naturalized populations in the British Isles (Roberts, 1964; Silverside, 1998; Stace et al., 2015). The *M. luteus* group includes a number of interfertile taxa: *M. luteus* var. *luteus*, *M. luteus* var. *variegatus*, *M. naiandinus*, *M. cupreus*, and *M. depressus* (Grant, 1924; Watson, 1989; von Bohlen, 1995; Cooley and Willis, 2009), some of which have become naturalized in the British Isles (Stace, 2010). Extant populations of *M. luteus* in the British Isles include highly polymorphic individuals, which may have resulted from hybridization events between different naturalized varieties. Here we focus on interspecific hybrids between *M. guttatus* and different varieties of *M. luteus* s.l., excluding hybrids with another closely related South American taxon, *M. cupreus*, which are rarer and easily distinguished (Stace, 2010). Hybrids between *M. guttatus* and *M. luteus* s.l. receive different taxonomic names, mostly based on color patterns of the corolla (Stace, 2010). For simplicity, hereafter we refer to these hybrids as *M. xrobertsii*, and the interested reader is referred to Silverside (1998) for a more detailed morphological classification of hybrid types.

Hybridization between diploid *M. guttatus* ( $2n = 2x = 28$ ) and tetraploid *M. luteus* ( $2n = 4x = 60$ – $62$ ) yields highly sterile, triploid individuals ( $2n = 3x = 44$ – $46$ , *M. xrobertsii*). Despite their sterility (Roberts, 1964; Parker, 1975), these triploid hybrids can be vegetatively vigorous and have a strong capacity to reproduce by clonal propagation from plant fragments that root at the nodes. *Mimulus robertsii* has been very successful at establishing naturalized populations throughout the British Isles (Stace et al., 2015), and it can

become locally abundant, forming populations of thousands of individuals. A recent genetic survey of *M. xrobertsii* showed that populations of this sterile hybrid are genetically variable and composed of multiple highly heterozygous genotypes (Vallejo-Marín and Lye, 2013). Approximately 40–50% of extant *Mimulus* populations in the United Kingdom contain *M. xrobertsii* (Vallejo-Marín and Lye, 2013).

**Crossing design**—To generate the plants used in this study, seeds were collected from natural populations of *M. guttatus* from Dunblane, Perthshire, Scotland (56.19°N, 3.98°W), and *M. luteus* s.l. from Coldstream, Scottish Borders, Scotland (55.65°N, 2.24°W) on 8 August 2009 and 10 July 2010, respectively. Individuals of *M. luteus* s.l. from Coldstream are polymorphic for the number of purple blotches present in the corolla lobes. Phenotypically, these *M. luteus* s.l. plants resemble intraspecific hybrids between *M. luteus* var. *luteus* and *M. luteus* var. *variegatus* (sometimes called *M. smithii*). From each population, we randomly chose six individuals collected as seeds from separate maternal plants and crossed them to generate 12 outbred lines for each species using an equal contribution design in which each individual contributed to the genome of four lines, two as the paternal parent and two as the maternal parent.

To generate interspecific hybrids, each of the 12 *M. guttatus* outbred lines were randomly paired with one of 12 *M. luteus* s.l. outbred lines. Two of these interspecific pairs were then assigned to each of six mating groups, and each of these groups was subject to a diallel cross, except selfs. To avoid half-sib matings, conspecifics in the same mating group were chosen to be unrelated. Therefore, all matings within each group are outbred and have the same inbreeding coefficient relative to the parental generation. Flowers were emasculated before conducting hand pollinations. Only 33 of 72 possible hybrid crosses were performed due to flower availability.

The crossing design yielded individuals of four different types: (1) *M. guttatus* × *M. guttatus* (MG × MG); (2) *M. luteus* s.l. × *M. luteus* s.l. (ML × ML); and reciprocal F1 hybrid crosses of (3) *M. guttatus* × *M. luteus* s.l. (MG × ML), and (4) *M. luteus* s.l. × *M. guttatus* (ML × MG). Hereafter, crosses are represented as maternal

× paternal parent. In total, we obtained seeds of 33 lines of the following crossing types: MG × MG (4 lines), ML × ML (10 lines), MG × ML (10 lines), and ML × MG (9 lines). For all subsequent experiments, we randomly chose four lines of each crossing type (16 lines total; Appendix S2; see online Supplemental Data).

**Seed production**—To determine the number of seeds produced per fruit, between 14 to 20 d after hand pollination, we selected 9 to 11 randomly selected fruits of each parental and F1 cross types were collected into glassine paper bags. The fruits were dissected using a stereomicroscope (MZ6, Leica Microsystems, Milton Keynes, UK) at 6.3× to 40× magnification to remove the fruit casing and count all seeds.

**Germination rate**—For determining seed viability and germination, for the F1 crosses, 100 randomly selected seeds from each line were surface-sown in 3.5 inches, 370 mL, round pots (9F; Desch Plantpak, Waalwijk, Netherlands) using a low nutrient compost (Modular Seed Growing Medium, William-Sinclair, Lincoln, UK). The pots were arranged in plastic trays in standing water (approximately 3–4 cm water depth). Trays were placed in controlled environment chambers (Microclima; Snijders Scientific B.V, Tilberg, Netherlands) at the University of Stirling with 16 h light (24°C) and 8 h dark (16°C) at 70% constant humidity. Seedling emergence was scored daily by counting the cumulative number of individuals in which the cotyledons could be observed.

**Floral and vegetative phenotyping**—For comparing the phenotypes of parental and hybrid lines, a second set of seeds from the same lines was planted and germinated as described above. Individual seedlings of each line were transplanted 10 d after germination into 3.5-inch pots with low-nutrient compost and arranged in flooded, plastic trays. Trays were placed in the University of Stirling glasshouses under natural light conditions, with 16 h of supplemental light provided by compact-fluorescent lamps (2700 K and 6400 K). Heating was only provided if the temperature inside the glasshouse dropped below 15°C during the day and 10°C at night. Plants were sprayed once to control for an aphid infestation, with 1.5 L of diluted Provado Ultimate Bug Killer (20 mL/L) (Bayer Garden, Cambridge, UK). This experiment was carried out between 3 July and 2 October 2012. The date and plant height at the time when the first flower opened was recorded, as well as the following measurements from the first two flowers of each plant using digital calipers: (1) corolla width (maximum distance between lateral petal lobes in frontal view); (2) corolla height (maximum distance between upper and lower petal lobes in frontal view); (3) corolla throat (size of the opening of the corolla tube near the center of the flower); (4) calyx length (measured at the upper sepal); (5) tube length (measured in side view); (6) pedicel length; (7) bract width (width of the subtending leaf at the node of the flower being measured); and (8) bract length (including petiole). Nectar production was measured on a subset of flowers using a 50 µL calibrated glass pipette (Drummond Scientific, Broomall, Pennsylvania, USA). The fertility of reciprocal crosses was not quantified here, but previous work (Roberts, 1964) and our preliminary observations suggest that triploid hybrids are sterile regardless of crossing direction.

**Confirming the success of artificial crosses**—To further check that the individuals produced through artificial crosses were indeed derived from interspecific hybridization, we genotyped all plants used

in the morphological analyses. Leaf tissue samples were collected from August to October 2012 and stored in silica gel. DNA was extracted using a modification of the CTAB protocol described by Doyle and Doyle (1990), and quantified with a Nanodrop 2000 spectrophotometer (Thermo Scientific, Wilmington, Delaware, USA). We genotyped putative hybrids and parents at six microsatellite loci (ATT240, AAT217, ATT267, AAT230, AAT225, AAT300; [Kelly and Willis, 1998]) following the protocol described by Vallejo-Marín and Lye (2013). We used the program STRand 2.4.59 (Toonen and Hughes, 2001) to analyze fluorescence profiles. Based on the genotyping results, four ML × MG individuals were identified as self-pollinations of the ML maternal plant based on lack of expected heterozygosity and excluded from subsequent analyses. While we do not have cytological data for individuals included in this experiment, we have confirmed the triploid nature of other putative triploid individuals in the same interspecific crosses using genome size estimates of nuclei stained with propidium iodide, and analyzed in a Guava easyCyte 5 flow cytometer (Merck Millipore, Watford, UK).

**Data analysis**—The number of seeds produced by the four cross types (MG × MG, MG × ML, ML × MG, and ML × ML) was statistically compared using a generalized linear model (with normal distribution) with cross type as the explanatory variable using the glm package in R ver. 3.1.2 (R Core Team, 2015). Pairwise post hoc comparisons among cross types was done using Tukey tests on the fitted model using multcomp (Hothorn et al., 2008). Germination curves for each cross type were calculated as the mean of the four studied lines. The proportion of plants that flowered was analyzed using a binomial distribution (logit link) in glm, and post hoc pairwise comparisons among cross types done with Tukey tests in multcomp. Nectar content and plant height were also analyzed using glm with cross type as an explanatory variable, and Tukey tests performed in multcomp. Floral and leaf traits are presented as the mean and standard error for illustration only, and no further generalized linear model or pairwise post hoc comparisons for cross types were done on these traits. A principal component analysis (PCA) of the 10 floral and vegetative traits shown in Table 4 was conducted using the function princomp on the correlation matrix.

**Chloroplast assembly**—Despite the availability of both nuclear (Hellsten et al., 2013) and mitochondrial (Mower et al., 2012) genomes for *M. guttatus*, no chloroplast genome for any *Mimulus* species had been previously published. Thus we sought to generate an assembly of the entire chloroplast genome using newly generated whole-genome sequences of an advanced inbred line of *M. luteus* var. *luteus* (MLl1/MLl2, Table 2). MLl1 and MLl2 are the same inbred line but are listed separately in Table 2 because they were genotyped via different methods (WGS and sequence capture). Whole-genome shotgun reads were trimmed using the program Trimmomatic v.0.32 (Bolger et al., 2014) with the SLIDINGWINDOW:10:20 and MINLEN:40 parameters. Trimmed reads were mapped against the *Salvia miltiorrhiza* chloroplast genome (NC\_020431.1; Qian et al., 2013) using the program Bowtie2 v.2.1.0 (Langmead and Salzberg, 2012) with the default “very-sensitive-local parameter” set. Mapped reads were assembled using the program SPAdes v.3.1.0 (Bankevich et al., 2012) with *k*-mer sizes of 55 and 87 under the “only-assembler” option. The resulting contigs were assembled with the full trimmed data set using the program afn (<https://bitbucket.org/afn/afn>) with the following parameters:



**TABLE 2.** Sample names and sampling locations for *Mimulus* accessions used in this study.

| Taxon                           | Population     | Sample ID         | Short name | Location                                     | Latitude | Longitude | Data type* | Accession   |
|---------------------------------|----------------|-------------------|------------|--|----------|-----------|------------|-------------|
| <i>M. guttatus</i> (2x)         | AYR            | 10-AYR-10         | AYR1       | Ayr, Ayrshire, Scotland                      | 55.461   | −4.625    | WGS        | SRR1462346  |
|                                 | AYR            | 10-AYR-18-1       | AYR2       | Ayr, Ayrshire, Scotland                      | 55.461   | −4.625    | SC         | dryad.gg273 |
|                                 | CER            | 10-CER-10         | CER1       | Cerrigydrudion, Wales                        | 53.006   | −3.549    | WGS        | SRR1475232  |
|                                 | CER            | 10-CER-25-1       | CER2       | Cerrigydrudion, Wales                        | 53.006   | −3.549    | SC         | dryad.gg273 |
|                                 | CRO            | 12-CRO-5          | CRO        | Crowan, Cornwall                             | 50.163   | −5.293    | WGS        | SRR1482405  |
|                                 | DBL            | 10-DBL-20         | DBL1       | Dunblane, Scotland                           | 56.197   | −3.965    | WGS        | SRR1475385  |
|                                 | DBL            | CG-1-1-JP1        | CG         | Dunblane, Scotland                           | 56.197   | −3.965    | WGS        | this study  |
|                                 | DBL            | 11-DBL-1-1        | DBL2       | Dunblane, Scotland                           | 56.187   | −3.965    | SC         | dryad.gg273 |
|                                 | HOU            | 10-HOU-17         | HOU1       | Houghton Lodge, England                      | 51.097   | −1.508    | WGS        | SRR1481643  |
|                                 | HOU            | 10-HOU-23-1       | HOU2       | Houghton Lodge, England                      | 51.097   | −1.508    | SC         | dryad.gg273 |
|                                 | PAC            | 12-PAC-39         | PAC        | Packhorse Bridge, Scotland                   | 57.355   | −3.336    | WGS        | SRR1482406  |
|                                 | QUA            | 10-QUA-47         | QUA        | Quarff, Shetland Islands                     | 60.105   | −1.227    | WGS        | SRR1481644  |
|                                 | TOM            | 10-TOM-23         | TOM        | Tomintoul, Scotland                          | 57.255   | −3.368    | WGS        | SRR1482404  |
|                                 | TRE            | 12-TRE-17         | TRE        | Tremar Coombe, England                       | 50.498   | −4.465    | WGS        | SRR1482409  |
|                                 | TSG            | TSG               | TSG        | Haida Gwaii, Canada                          | 53.419   | −131.916  | WGS        | SRS120816   |
|                                 | VIC            | 12-VIC-18         | VIC        | Victoria Bridge, Northern Ireland            | 54.763   | −7.454    | WGS        | SRR1482407  |
|                                 | AHQ            | AHQ               | AHQ        | Lonestar Basin Thermal Spring, Teton Co., WY | 44.431   | −110.813  | WGS        | SRR486613   |
|                                 | BOG            | BOG               | BOG        | Bog Hot Springs, Humboldt, NV                | 41.924   | −118.804  | WGS        | SRR071996   |
|                                 | LMC            | LMC               | LMC        | Yorkville, Mendocino Co., CA                 | 38.864   | −123.084  | WGS        | SRR072031   |
|                                 | MAR            | MAR               | MAR        | Marshanne Landing, Douglas Co., OR           | 43.479   | −123.294  | WGS        | SRR071967   |
|                                 | PED            | PED               | PED        | San Pedro River, Pinal Co., AZ               | 32.711   | −110.628  | WGS        | SRR071969   |
|                                 | REM            | REM               | REM        | Rumsey, Yolo Co., CA                         | 38.859   | −122.410  | WGS        | SRR071971   |
|                                 | SWB            | SWB               | SWB        | Sperm Whale Beach, Mendocino Co., CA         | 39.036   | −123.691  | WGS        | SRR072030   |
|                                 | YJS            | YJS               | YJS        | Yellowjacket creek, Lemhi Co., ID            | 44.951   | −114.585  | WGS        | SRR071970   |
| <i>M. guttatus</i> complex      |                |                   |            |  |          |           |            |             |
| <i>M. platycalyx</i>            | CVP            | CVP               | CVP        | Coleman Valley Road, Sonoma Co., CA          | 38.372   | −123.055  | WGS        | SRR486607   |
| <i>M. micranthus</i>            | EBR            | EBR               | EBR        | Branscomb, Mendocino Co., CA                 | 39.631   | −123.532  | WGS        | SRR486609   |
| <i>M. nasutus</i>               | SF             | SF                | SF         | Sherars Falls, Wasco Co., OR                 | 45.635   | −120.914  | WGS        | SRR400478   |
| <i>M. luteus</i> s.l. (4x)      | COL            | CS-4-JP1          | CS         | Coldstream, Scotland                         | 55.655   | −2.240    | WGS        | this study  |
|                                 | COL            | 10-COL-20e-1      | COL        | Coldstream, Scotland                         | 55.655   | −2.240    | SC         | dryad.gg273 |
|                                 | EVI            | 13-EVI-5          | EVI        | Evie, Scotland                               | 59.112   | −3.108    | SC         | dryad.gg273 |
| <i>M. l. luteus</i>             | EY             | MLI-EY            | MLI1       | El Yeso, Chile                               | −33.4    | −70.0     | WGS        | this study  |
| <i>M. l. luteus</i>             | EY             | MLI-EY-7x4-1-3    | MLI2       | El Yeso, Chile                               | −33.4    | −70.0     | SC         | dryad.gg273 |
| <i>M. l. variegatus</i>         | RC             | MLv-RC-5x4-4-1    | MLv        | Río Cipreses, Chile                          | −34.2    | −70.3     | SC         | dryad.gg273 |
| <i>M. robertsii</i> (3x)        | DBL × COL (3x) | H003b-JP1         | HOO        | Artificial triploid (CG-1-1 × CS-4-3)        |          |           | WGS        | this study  |
|                                 | GIO            | 13-GIO-13         | GIO        | Giordail Beach, Scotland                     | 58.337   | −6.202    | SC         | dryad.gg273 |
|                                 | LED            | 12-LED-3-11       | LED1       | Leadhills, Scotland                          | 55.424   | −3.735    | SC         | dryad.gg273 |
|                                 | NEN            | 12-NEN-41         | NEN        | Nenthall, England                            | 54.806   | −2.376    | SC         | dryad.gg273 |
| <i>M. peregrinus</i> (6x)       | TOR            | 13-TOR-13         | TOR        | Tormiston, Scotland                          | 58.996   | −3.183    | SC         | dryad.gg273 |
|                                 | DBL × COL (6x) | 13-SYN-1-JP1      | SYN        | Artificial hexaploid (CG-1-1 × CS-4-3)       |          |           | WGS        | this study  |
|                                 | GON            | 12-GON-106        | GON        | Glen Gonnar, Scotland                        | 55.467   | 3.738     | SC         | dryad.gg273 |
|                                 | LED            | 11-LED-seed-1-JP1 | LED2       | Leadhills, Scotland                          | 55.423   | −3.735    | WGS        | this study  |
|                                 | LED            | 11-LED-seed-30    | LED3       | Leadhills, Scotland                          | 55.423   | −3.735    | SC         | dryad.gg273 |
|                                 | LED            | 11-LED-seed-2-11  | LED4       | Leadhills, Scotland                          | 55.424   | −3.735    | SC         | dryad.gg273 |
| <i>M. dentilobus</i> (Outgroup) | STR            | 13-STR-32         | STR        | Stromness, Scotland                          | 58.969   | −3.283    | SC         | dryad.gg273 |
|                                 | DENT           | DENT              | DENT       | NA   | NA       | NA        | WGS        | SRX030541   |

Notes: WGS, whole-genome sequencing; SC, sequence capture.

a stop extension value of 0.1, an initial trim of 100 bp to contigs, a maximum extension of 100 bp per loop, and 50 search loops. *afin* spans contigs by trimming contig ends, identifying matching reads to the contig ends, and extending the contigs iteratively while attempting to fuse contigs at each iteration. Ultimately, *afin* was able to assemble the plastome from ~10 contigs to one. The assembled plastome was put into Sequencher v. 5.0.1 (Gene Codes Corp., Ann Arbor, Michigan, USA), and the boundaries of the quadripartite regions (large single copy-LSC, inverted repeat B-IRB, small single copy-SSC, and inverted repeat A-IRA) were identified. The final plastome is presented in the LSC-IRB-SSC-IRA format. The final assembly of the plastome was verified through a coverage analysis. The program Jellyfish v.2.1.3 (Marçais and Kingsford, 2011) was

used to estimate 25-mer abundance from the cleaned reads. These abundances were used to map a 25-bp sliding window of coverage across the assembled plastome. The coverage was found to be uniform and equal for the single copy regions (LSC = 373×; SSC = 328×) with the inverted repeat having ~2× the coverage of the single copy regions (IR = 756×), as expected. The program DOGMA (Wyman et al., 2004) was used to annotate the *Mimulus luteus* plastome. A graph of the annotated plastome structure was created using the program Circos v. 0.66 (Krzywinski et al., 2009), as implemented in the Verdant database ([verdant.iplantcollaborative.org](http://verdant.iplantcollaborative.org)).

**Samples for genomic parentage analysis**—To determine the relationships between the cytoplasmic genomes of *M. guttatus* and

*M. luteus*, we analyzed previously published and newly generated sequencing data of both parental taxa (Table 2). We included 24 individuals of *M. guttatus* from 10 introduced (UK) and nine native populations. Within the introduced range, we included replicate individuals for four populations (AYR, CER, DBL, and HOU) and a controlled cross of one of these populations (DBL). For *M. luteus*, we included an inbred line of *M. luteus* var. *luteus* and one inbred line of *M. luteus* var. *variegatus* both originally collected in the native range (Chile). The *M. luteus* var. *luteus* inbred line was represented by two individuals (ML1 and ML2, from the 7th and 9th inbred generations). In addition, we analyzed *M. luteus* s.l. from two populations in the introduced range in the United Kingdom (COL and EVI), including two individuals originating from the same population (COL): a wild collected plant and an individual from an experimental cross (Table 2). For *M. xrobertsii*, we included samples from four natural populations (GIO, LED, NEN, and TOR). For *M. peregrinus*, we analyzed samples from three populations (GON, STR, and LED), including three individuals from the type populations for the allopolyploid species (LED). As a control, we also analyzed an experimentally produced "*M. xrobertsii*" hybrid obtained by crossing *M. guttatus* × *M. luteus* s.l. (DBL × COL, maternal × paternal parent; CG-1-1 × CS-4-3 = H003b), as well as an S1 synthetic "*M. peregrinus*" allohexaploid produced by colchicine treatment of the H003b "*M. xrobertsii*" line (13-SYN-1-JP1). In addition, we also included three samples from taxa closely related to *M. guttatus* in North America (*M. nasutus*, *M. micranthus*, and *M. platycalyx*), plus one outgroup (*M. dentilobus*) (Table 2). In terms of sequencing approach, 16 samples (four of each *M. guttatus*, *M. luteus*, *M. xrobertsii*, and *M. peregrinus*) were derived using sequence capture (Vallejo-Marín et al., 2015), and the remaining through whole-genome sequencing obtained from previously published data (10 samples; Puzey and Vallejo-Marín, 2014, 13 samples from the Sequence Read Archive (deposited by the Joint Genome Institute; <http://www.ncbi.nlm.nih.gov/sra>; National Institutes of Health, Bethesda, Maryland, USA), or data newly generated here (six samples). In total, we analyzed 45 genomic samples (Table 2).

**Genotyping of plastid genomes**—We used the publicly available *M. guttatus* mitochondrial genome (Mower et al., 2012) (JN098455) or the newly assembled *M. luteus* var. *luteus* chloroplast genome (this study) as references for base calling and genotyping. All reads were aligned to the references using Bowtie2 (Langmead and Salzberg, 2012). Alignments were sorted, duplicates marked and discarded, and read groups added using Picard tools (<http://broadinstitute.github.io/picard>; Broad Institute, Cambridge, Massachusetts, USA). Following alignment filtration, genotypes were called using GATK UnifiedGenotyper (McKenna et al., 2010; DePristo et al., 2011; Auwera et al., 2013) (online Appendix S3). Following genotyping, indels and heterozygous sites were excluded.

**Filtering of genomic data for phylogenetic analysis**—The 17 chloroplast-derived regions of the *Mimulus guttatus* mitochondrial genome (NC\_018041.1; Mower et al., 2012) were extracted and aligned against the *Mimulus luteus* chloroplast genome using Sequencher v. 5.0.1 (Gene Codes). The *Mimulus luteus* chloroplast sequence was trimmed before alignment so that only the LSC, IRB, and SSC were present. Alignments were inspected for sequence similarity. Multiple instances where portions of the mitochondrial chloroplast-derived sequences did not have an ordinal alignment

with the majority of the sequence were identified. In these cases, the region that did not align was removed and realigned to the full plastome. These chimeric chloroplast-derived sequences may reflect multiple recombination events and subsequent recombination hotspots of the mitochondrial genome, or they are representative of multiple chloroplast-genome insertion events in the history of the mitochondrial genome. Once aligned, the coordinates of alignment to the chloroplast genome were recorded. Data corresponding to these positions were eventually removed from the data matrix (see below).

Chloroplast sequences were reconstructed from VCF files representing mapping to the *Mimulus luteus* plastome (see above). For each position of the *M. luteus* plastome, the corresponding base in the mapped data set was recorded. If an indel was identified, that position was ignored. Indels are difficult to distinguish from missing data using this method, so we chose to completely ignore them to reduce potentially conflicting signal. Each reconstructed plastome was split into features (protein-coding genes, tRNAs, and rRNAs) and interfeature regions, which were used for individual alignments and concatenation (see below). Features were identified based on their alignment to known coordinates of the annotated *Mimulus luteus* chloroplast genomes. Chloroplast regions identified in mitochondrial genome were excluded in phylogenetic analyses. This process was done using a Perl script ([https://github.com/mrmckain/AJB\\_2015\\_Mimulus\\_Allopolyploidy](https://github.com/mrmckain/AJB_2015_Mimulus_Allopolyploidy)).

Mitochondrial sequences were reconstructed in a similar manner to that of the plastid sequences. In this case, the annotation of the *M. guttatus* mitochondrial genomes was used to recognize features in the reconstructed sequences. Regions mapping to the chloroplast-derived regions of the mitochondrial were ignored. As before, reconstructed features and interfeature regions were used in alignments and then concatenated (see below).

**Phylogenetic analyses**—Individual features and interfeature regions were aligned using the program MAFFT v. 7.029b (Katoh et al., 2005) on auto settings. Individual alignments were concatenated into a single alignment that was assessed for presence/absence of nucleotides at each site. For the chloroplast alignment, if a gap was identified at a site, that site was removed. As with indels, we chose to do this to alleviate the effects of missing data on the phylogeny estimation. The final, ungapped alignment was 86,843 bp for all 45 samples. The chloroplast phylogeny was estimated using the program RAxML v. 8.0.22 (Stamatakis, 2006) under the GTR+Γ model of evolution with 500 bootstrap replicates.

Mitochondrial sequences were processed in a similar fashion to the plastid sequences. The individual feature and interfeature regions were aligned using MAFFT and concatenated. The mitochondrial sequence data were missing a larger percentage of the total mitochondrial genome, on average, across the 45 sample set. We noted that the sequence-captured samples were extremely low in their total representation of the mitochondrial genome. Filtering out all gapped sites (as above) resulted in a very short alignment of 4875 bp. After removal of the sequence-captured samples and the WGS sample CG, which demonstrated a low coverage of the mitochondrial genome, the ungapped alignment of the remaining 28 accessions totaled 402,717 bp. The mitochondrial phylogeny was estimated as above using the reduced 28 sample set.

We included CS-4-3 (*M. luteus*), CG-1-1 (*M. guttatus*), first generation laboratory-generated triploid hybrid H003b (CG-1-1 × CS-4-3), and laboratory-generated allohexaploid SYN (13-SYN-1-JP1) to

verify that this approach could determine maternal ancestry in a controlled hybridization event. In addition, we also included two *M. luteus* individuals (ML1 and ML2) derived from the same inbred line but genotyped using different methods (WGS and SC, respectively). Alignments and phylogenetic trees are available in the Dryad Data Repository (doi:10.5061/dryad.5q91d).

**Haplotype networks**—We estimated matrilineal haplotype networks for both chloroplast and mitochondria using statistical parsimony (TCS) implemented in the program PopArt (<http://popart.otago.ac.nz>). For this analysis, we used a panel of 929 SNPs for the chloroplast and 1454 for the mitochondria. Only variable sites were included in the SNP matrix. Indels and sites filtered as described in the section *Filtering of genomic data for phylogenetic analysis* were excluded. For chloroplast, we included both whole-genome sequence (WGS) and sequence capture (SC) samples, and for mitochondria we excluded the SC samples as above. For both data sets, we excluded sites with more than 5% missing data, allowing us to analyze 493 sites of which 161 were parsimony-informative for chloroplast and 784 sites of which 300 were parsimony-informative for the mitochondria.

**Principal component analysis**—We also conducted a PCA using the SNP data of the chloroplast and mitochondrial genomes using the glPca function in the R package adegenet v. 2.0 (Jombart and Ahmed, 2011). For this analysis, we removed genomic regions with low genotyping success and included only those SNP sites that were successfully genotyped in 90% (41/45) or more individuals. This filtering step removed poorly genotyped genomic regions and reduced the total number of SNPs from 1454 to 434 for the mitochondrial genome and from 929 to 694 in the chloroplast. For the chloroplast analysis, we also removed a single WGS individual that had poor genotyping success across the genome even after filtering (CG). As expected the amount of missing data across genotyped SNP loci was higher for the SC than for the WGS data set after filtering. For the mitochondrial data set, the amount of missing data per site per individuals was 0.0013 for WGS and 0.1080 for SC. For the chloroplast, the same threshold yielded an average amount of missing data per site per individuals of 0.0054 for WGS and 0.0176 for SC.

**Mitochondrial genotyping in natural hybrids**—We used mitochondrial genome sequences to identify potential markers that could help us distinguish *M. guttatus* and *M. luteus* in natural hybrid populations. We sought loci segregating for alternative alleles in *M. luteus* (ML1 and CS) and *M. guttatus* (CG and AYR1) in the sequenced panel. Our goal was to identify fragment length polymorphic sites (indels) that could be scored using microcapillary fragment analysis. To identify segregating indels, we sorted BAM alignments in IGV genome browser, and manually screened for indels that were alternatively fixed between *M. luteus* and *M. guttatus*. We identified three indels (3–4 bp) that distinguished *M. guttatus* and *M. luteus*. Genotyping individuals at these indels produced a distinct haplotype with which we could potentially identify the maternal (mitochondrial) parent of hybrid individuals. To facilitate large-scale genotyping, we designed primers flanking each of the indel regions using the program Primer3Plus (Untergasser et al., 2012) and confirmed their amplification and sequence in a separate test panel of parental and hybrid individuals using Sanger sequencing. We checked the compatibility of the three primer pairs for

pooling them in a single multiplex reaction using the program Multiplex Manager v 1.2 (Holleley and Geerts, 2009). We then used fluorescently labeled forward primers (6-FAM, Eurofins Genomics, Ebersberg, Germany), and unlabeled reverse primers to generate PCR products in a single multiplex reaction (Type-It, Qiagen, Manchester, UK). PCR cycles consisted of a denaturing step of 5 min at 95°C; followed by 30 cycles of 95°C for 30 s, 55°C for 180 s, and 72°C for 30 s; and a final elongation step of 30 min at 60°C. Fragment length of the PCR products was measured in an ABI3730 sequencer (Applied Biosystems, Carlsbad, California, USA) using LIZ500 as a size standard (DNA Sequencing and Services, Dundee, UK). Primer sequences for these mitochondrial markers are given in online Appendix S4. We genotyped 163 individuals of *M. guttatus* (3 populations), *M. luteus* s.l. (4 populations), *M. xrobertsii* (13 populations), and *M. peregrinus* (2 populations), and a synthetic allohexaploid line product of an *M. guttatus* (DBL) × *M. luteus* (COL) cross, treated with colchicine and selfed for one generation (SYN; Table 2). Between 1 and 15 individuals were genotyped per taxon per population ( $6.8 \pm 3.5$ , mean  $\pm$  SD) (Table 3).

## RESULTS

**Seed set and germination of artificial hybrids**—Reciprocal crosses between *M. guttatus* and *M. luteus* resulted in a relatively high number of seeds per fruit in both directions. The mean number ( $\pm$ SE) of seeds per fruit was  $723 \pm 73$  for the MG × ML cross, and  $635 \pm 100$  for ML × MG. This difference was not statistically significant (Tukey test,  $P = 0.72$ ; Fig. 1A). Within-species crossing yielded  $1237 \pm 95$ , and  $467 \pm 57$  seeds per fruit for *M. guttatus* and *M. luteus*, respectively (Fig. 1A). In comparison, we found a marked effect of the identity of the maternal parent on seed germination. While approximately half of the seeds of the MG × ML cross germinated ( $0.525 \pm 0.062$ ), only 4% of the ML × MG cross were able to do so ( $0.042 \pm 0.018$ ) (Fig. 1B). Both parental taxa had high germination proportions ( $0.835 \pm 0.037$  and  $0.825 \pm 0.020$ , for MG × MG, and ML × ML, respectively), although *M. guttatus* seeds achieved maximum germination at a much faster rate (Fig. 1B).

**Floral and vegetative phenotype of F1 hybrids**—The proportion of hybrid plants that flowered was significantly different depending on crossing direction. The MG × ML cross had a higher proportion of flowering individuals ( $0.983 \pm 0.017$ ,  $n = 60$ ) than the reciprocal cross did ( $0.706 \pm 0.114$ ,  $n = 17$ ) (Tukey test  $z = 2.809$ ,  $P = 0.021$ ). The proportion of flowering plants for the MG × ML cross also exceeded both parental crosses ( $0.633 \pm 0.044$ ,  $n = 120$ ; and  $0.741 \pm 0.040$ ,  $n = 120$ ; for MG × MG [ $z = 3.442$ ,  $P < 0.01$ ] and ML × ML [ $z = 2.979$ ,  $P < 0.05$ ], respectively). Moreover, MG × ML hybrids began flowering on average 9 d earlier than the reciprocal cross did ( $54.91 \pm 0.96$  d,  $63.25 \pm 4.08$  d). Flowering occurred slightly earlier for *M. luteus* ( $53.08 \pm 0.98$  d) than for *M. guttatus* ( $57.77 \pm 1.14$  d).

Among plants that flowered, the MG × ML hybrid achieved the largest stature of all cross types ( $267 \pm 9$  mm vs.  $139.1 \pm 12.3$ ,  $188.7 \pm 5.0$ , and  $138.9 \pm 5.4$ ; for ML × MG [ $z = 7.491$ ,  $P < 0.001$ ], MG × MG [ $z = 8.431$ ,  $P < 0.001$ ], and ML × ML [ $z = 14.011$ ,  $P < 0.001$ ], respectively; Fig. 2). Flowers and bracts of both types of hybrids were intermediate or larger than the parental values (Table 4; online Appendices S5, S6). Both MG × ML and ML × MG hybrids had similar-sized flowers, although hybrids with *M. guttatus* as the maternal parent had longer pedicels and larger bracts (Table 4).



**TABLE 3.** Mitochondrial haplotypes of 163 individuals of *M. guttatus*, *M. luteus*, and their hybrid derivatives, the triploid *M. xrobertsii* and allopolyploid *M. peregrinus*. Haplotypes are given as the size in base pairs of three mitochondrial loci with fragment-length polymorphism (indels). Between 1 and 15 individuals of each of 21 populations, and one artificially created allopolyploid, were genotyped at all three loci. The synthetic allopolyploid was the product of a *M. guttatus* (DBL) × *M. luteus* s.l. (COL) cross, with *M. guttatus* as the maternal parent. A single individual phenotyped as *M. guttatus* (DBL), but with a “*M. luteus*” maternal haplotype is presented separately. Individuals with the *M. luteus* haplotype are in boldface.

| Taxon                                   | Population            | No. samples | Haplotype          | Inferred maternal parent |
|---|-----------------------|-------------|--------------------|--------------------------|
| <i>M. guttatus</i>                      | COL                   | 5           | 237/296/313        | <i>M. guttatus</i>       |
|   | DBL                   | 5           | 237/296/313        | <i>M. guttatus</i>       |
|   | DBL                   | 1           | <b>240/292/317</b> | <b><i>M. luteus</i></b>  |
|   | TOM                   | 3           | 237/296/313        | <i>M. guttatus</i>       |
| <i>M. luteus</i> s.l.                   | COL                   | 5           | <b>240/292/317</b> | <b><i>M. luteus</i></b>  |
|   | EVI                   | 3           | <b>240/292/317</b> | <b><i>M. luteus</i></b>  |
| <i>M. luteus</i> var. <i>luteus</i>     | EY, Chile             | 1           | <b>240/292/317</b> | <b><i>M. luteus</i></b>  |
| <i>M. luteus</i> var. <i>variegatus</i> | RC, Chile             | 1           | <b>240/292/317</b> | <b><i>M. luteus</i></b>  |
| <i>M. xrobertsii</i>                    | BCH                   | 6           | 237/296/313        | <i>M. guttatus</i>       |
|   | EOR                   | 9           | 237/296/313        | <i>M. guttatus</i>       |
|   | EVI                   | 1           | 237/296/313        | <i>M. guttatus</i>       |
|   | GIO                   | 8           | 237/296/313        | <i>M. guttatus</i>       |
|   | GON                   | 9           | 237/296/313        | <i>M. guttatus</i>       |
|   | GOO                   | 10          | 237/296/313        | <i>M. guttatus</i>       |
|   | GRO                   | 10          | 237/296/313        | <i>M. guttatus</i>       |
|   | NEN                   | 9           | 237/296/313        | <i>M. guttatus</i>       |
|   | NEN                   | 3           | <b>240/292/317</b> | <b><i>M. luteus</i></b>  |
|   | NOR                   | 9           | 237/296/313        | <i>M. guttatus</i>       |
|   | PLY                   | 7           | 237/296/313        | <i>M. guttatus</i>       |
|   | SCA                   | 9           | 237/296/313        | <i>M. guttatus</i>       |
|   | TAR                   | 10          | 237/296/313        | <i>M. guttatus</i>       |
|   | TOR                   | 10          | 237/296/313        | <i>M. guttatus</i>       |
|   | LED                   | 15          | 237/296/313        | <i>M. guttatus</i>       |
| <i>M. peregrinus</i>                    | STR                   | 8           | 237/296/313        | <i>M. guttatus</i>       |
|   | Synthetic (DBL × COL) | 6           | 237/296/313        | <i>M. guttatus</i>       |
| Total                                   | 22                    | 163         |                    |                          |

Parental taxa differed significantly in nectar production (ML × ML produced more nectar than MG × MG did; Tukey test  $z = 4.998$ ,  $P < 0.001$ ), but while MG × ML hybrids produced significantly more nectar than MG × MG ( $z = 6.138$ ,  $P < 0.001$ ), ML × MG hybrids produced larger amounts, but not significantly so ( $z = 2.299$ ,  $P = 0.092$ ), than the *M. guttatus* parent (Fig. 2).

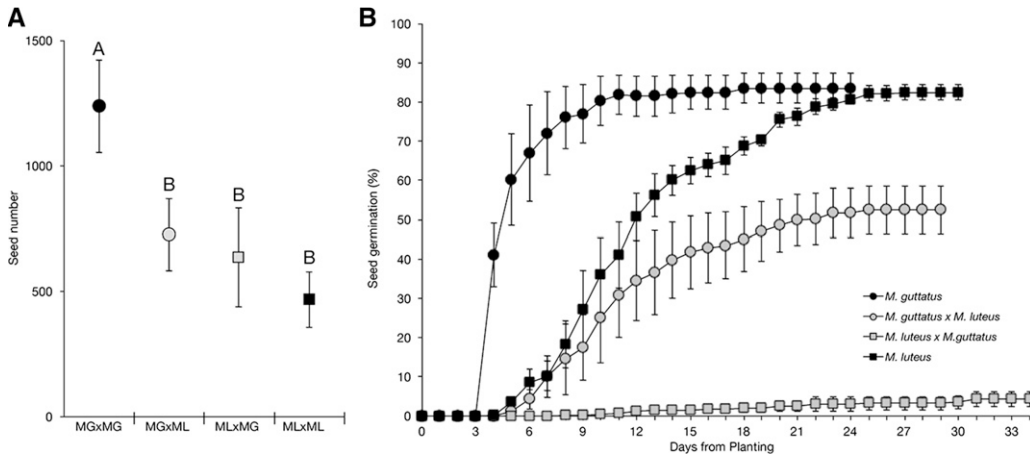
**Analysis of *Mimulus* chloroplast**—The complete *Mimulus luteus* chloroplast genome was assembled, providing the first published sequence for both the genus *Mimulus* and the family Phrymaceae. The plastome follows the typical quadripartite structure found in most of the photosynthetic angiosperm lineages. The total length of the chloroplast genome is 153,150 bp and is made up of a 84,293 bp LSC, a 17,851 bp SSC, and two 25,503 bp IR regions (Fig. 3). The complete chloroplast sequence is available in GenBank (KU705476). The plastome contains 80 unique protein-coding genes (including *ycf1*, *ycf15*, and *ycf2*), 30 unique tRNAs, and 4 unique rRNAs. No major gene losses or rearrangements were identified relative other published plastomes from Lamiales (Qian et al., 2013; Zhang et al., 2013).

**Genomic analysis of hybrid parentage based on cytoplasmic genomes**—On average, sequencing coverage of the cytoplasmic genomes was very high, although also very skewed. For WGS data sets, per individual average read coverage of the mitochondrial and chloroplast genome ranged from 48× to 940× and 347× to 5556×

with global means of 318 and 2303, respectively. As expected, the sequence capture data sets had high read coverage, but totals were considerably lower than the WGS data sets (Appendix S2). Per individual average read coverage for SC data sets mapped against the mitochondrial and chloroplasts genomes from 6× to 13× and 55× to 112× with means of 9× and 85×, respectively. The chloroplast had considerably greater read depth than the mitochondria did. The fraction of genotyped bases varied considerably between the WGS and SC data set (online Appendix S7). On average, 68% and 80% of chloroplast sites were genotyped in the SC and WGS data sets, respectively (fraction of sites genotyped calculated from entire mitochondrial and chloroplast data sets before filtering). The mitochondria had a larger span between the average percentage of called bases between the SC and WGS data sets with 57% and 85% called.

The chloroplast tree had relatively high resolution and strong support at the taxon level (i.e., *M. guttatus* vs. *M. luteus*), but was less useful to resolve relationships between populations within species (Fig. 4), suggesting recent divergence of populations. A single monophyletic *M. luteus* clade (minus MLv) is well supported in the chloroplast tree. The mitochondrial data set does not support *M. luteus* monophyly. Within the chloroplast tree, *M. peregrinus* (allopolyploid) forms two distinct groups that are more closely related to geographically proximate sterile triploid hybrids. *Mimulus xrobertsii* sample (LED1) and *M. peregrinus*, LED2, LED3, and LED4 are closely related, and were all collected from Leadhills, Scotland. *Mimulus peregrinus* sample GON also falls within the above group and is from nearby Glengonnar, Scotland. Based on the chloroplast data, the other naturally *M. peregrinus* sample included in this analysis, STR, also falls within a clade with geographically proximate *M. xrobertsii* (TOR). The relationship of *M. xrobertsii* and *M. peregrinus* supports the finding of Vallejo-Marín et al. (2015) using a different aspect of the genomic data, i.e., the nuclear component of the genome. Naturally occurring allohexaploids (LED2, LED3, LED4, GON, and STR) as well as naturally occurring triploid hybrids (LED1, TOR, NEN, GIO) are more closely related to *M. guttatus* than they are to *M. luteus* based on chloroplast analysis. The mitochondrial tree topology indicates that naturally occurring allopolyploid LED is nested within a *M. guttatus* clade. The chloroplast tree places ML1 and ML2 (samples derived from the same inbred lines but genotyped using different methods) sister to each other. HOO and SYN, triploid hybrid and allohexaploid samples, respectively, derived from an artificial cross between CG × CS, fall within the same clade in both the chloroplast and mitochondrial trees. The maternal parent of these lines (CG) is not sister to the SYN and HOO samples as expected, most likely resulting from homoplasmy affecting tree reconstruction due to relatively low sequence divergence between accessions.



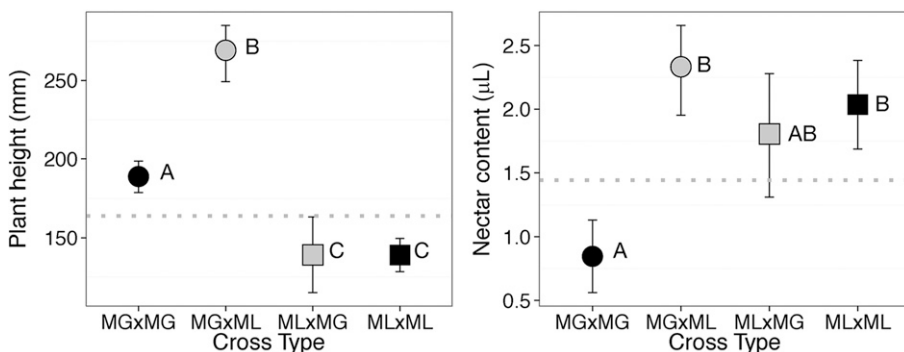


**FIGURE 1** (A) Mean seed number and 95% confidence intervals (CI) obtained in crosses within species (MG  $\times$  MG and ML  $\times$  ML), and between species (MG  $\times$  ML and ML  $\times$  MG). Sample size (fruits): 9, 11, 10, 10. Different letters indicate significantly different means as assessed with a pairwise Tukey test ( $z > 2.98$ ,  $P < 0.05$ ). (B) Mean germination of *M. guttatus*  $\times$  *M. guttatus* (black circle), *M. luteus*  $\times$  *M. luteus* (black square), *M. guttatus*  $\times$  *M. luteus* (gray circle), *M. luteus*  $\times$  *M. guttatus* (gray square). Crosses are indicated as maternal  $\times$  paternal parent. Four lines for each cross type were analyzed, planting 100 seeds per line (1600 seeds total). Each data point represents the average ( $\pm$ SE) proportion of germinated seeds across the four lines for a given cross type across the germination period.

The haplotype network of the chloroplast showed that both *M. xrobertsii* and *M. peregrinus* samples have identical or nearly identical haplotypes to *M. guttatus* (Fig. 5A). Samples of *M. peregrinus* from Leadhills (LED2, LED3, LED4) and nearby Glengonnar (GON) share a haplotype with *M. xrobertsii* from Leadhills (LED1). Samples of *M. peregrinus* from Orkney (STR) belong to a closely related haplotype that also includes *M. xrobertsii* from Orkney (TOR), northern England (NEN), and the Outer Hebrides (GIO). This haplotype also includes most of the *M. guttatus* samples from the British Isles, the synthetic triploid and hexaploid (HOO and SYN), and one native population (AHQT). Samples of *M. luteus* s.l. from the British Isles (CS and COL, Coldstream; EVI, Evie) fall in a distinct group of haplotypes separated from other *Mimulus* by several mutational steps and which also includes the two replicates from the *M. luteus* var. *luteus* inbred line (ML11 and ML12; Chile). As in the phylogenetic reconstruction, the sample from a Chilean *M. luteus* var. *variegatus* (MLv) is situated near *M. guttatus* haplotypes.

6A; online Appendix S8). In the mitochondrial PCA, there is not enough resolution to differentiate these two origins, and all *M. peregrinus* and *M. xrobertsii* samples fall in the same cluster, along with other native and introduced *M. guttatus* (Fig. 6B; online Appendix S9). As above, the *M. luteus* var. *variegatus* sample from Chile (MLv) is closer to *M. guttatus* samples than to other *M. luteus*.

**Natural frequency of *M. guttatus* as maternal parent**—We successfully genotyped 163 *Mimulus* spp. individuals from 21 populations and one synthetic line at three mitochondrial loci (Table 3). We recovered only two haplotypes: 237/296/313 and 240/292/317 (fragment size in base pairs for markers mit-259, mit-262, and mit-357, respectively). For *M. luteus*, we genotyped 10 individuals, all of which had haplotype 240/292/317. For *M. guttatus*, we genotyped 14 individuals from three populations of which all but one had haplotype 237/296/313. The exceptional individual belonged to population DBL (Table 3). We genotyped 110 individuals of *M. xrobertsii* from 12 populations across the UK. From these, 107 had haplotype 237/296/313 (“*M. guttatus* haplotype”), while three individuals from population NEN had the other haplotype. Finally, the 23 individuals of *M. peregrinus* sampled in populations STR and LED also had the haplotype 237/296/313. The synthetic allohexaploid line created with *M. guttatus* as the maternal parent, had haplotype 237/296/313 as expected.



**FIGURE 2** (A) Plant height at first flower and (B) nectar content per flower of *Mimulus* crosses within (MG  $\times$  MG and ML  $\times$  ML) and between species (MG  $\times$  ML and ML  $\times$  MG). Sample size (individuals) for plant height:  $n = 72, 54, 11, 78$ ; sample size (flowers):  $n = 67, 73, 14, 73$ . Dotted line shows the midparent value. Bars are 95% confidence intervals.

The mitochondrial network, which excludes all SC samples due to low genotyping success, shows the single individual of wild *M. peregrinus* (LED2) as part of a haplotype that includes most of the British *M. guttatus* as well as the synthetic triploid and hexaploid accessions (HOO and SYN). Both *M. luteus* var. *luteus* (ML11) and *M. luteus* s.l. (CS) are separated from all other *Mimulus* by a large number of mutational steps (Fig. 5B).

The PCA of both chloroplast and mitochondria data sets also showed that *M. xrobertsii* and *M. peregrinus* cluster together with *M. guttatus* samples (Fig. 6). In the chloroplast PCA, the two known origins of *M. peregrinus* (Leadhills and Orkney) fall in separate clusters with *M. guttatus*, each of them associated with local *M. xrobertsii* samples (Fig.

## DISCUSSION

Our study combines experimental crosses, genomic analyses of chloroplast and mitochondrial genomes, and surveys of natural hybrid populations to demonstrate how a

**TABLE 4.** Phenotypic measurements (mean  $\pm$  SE) of *M. guttatus* (MG  $\times$  MG), *M. luteus* (ML  $\times$  ML), and their reciprocal hybrids (MG  $\times$  ML, and ML  $\times$  MG). Crosses are shown as maternal  $\times$  paternal parent.

| Traits                 | MG $\times$ MG (n = 146) | MG $\times$ ML (n = 117) | ML $\times$ MG (n = 23) | ML $\times$ ML (n = 155) |
|------------------------|--------------------------|--------------------------|-------------------------|--------------------------|
| Floral traits (mm)     |                          |                          |                         |                          |
| Corolla height         | 23.54 $\pm$ 0.47         | 26.85 $\pm$ 0.44         | 27.43 $\pm$ 1.47        | 21.18 $\pm$ 0.42         |
| Corolla width          | 27.26 $\pm$ 0.4          | 30.53 $\pm$ 0.39         | 29.79 $\pm$ 1.44        | 24.21 $\pm$ 0.46         |
| Corolla throat         | 2.56 $\pm$ 0.07          | 3.63 $\pm$ 0.07          | 3.98 $\pm$ 0.26         | 4.43 $\pm$ 0.08          |
| Corolla tube length    | 20.74 $\pm$ 0.22         | 24.38 $\pm$ 0.24         | 24.45 $\pm$ 0.5         | 23.51 $\pm$ 0.27         |
| Flower depth           | 36.04 $\pm$ 0.47         | 36.54 $\pm$ 0.38         | 35.34 $\pm$ 1.15        | 31.53 $\pm$ 0.4          |
| Calyx length           | 13.12 $\pm$ 0.14         | 14.81 $\pm$ 0.15         | 15.03 $\pm$ 0.32        | 13.87 $\pm$ 0.12         |
| Pedicle length         | 18.91 $\pm$ 0.45         | 42.16 $\pm$ 1.31         | 35.31 $\pm$ 3.12        | 46.53 $\pm$ 1.01         |
| Vegetative traits (mm) |                          |                          |                         |                          |
| Bract length           | 16.58 $\pm$ 0.48         | 56.59 $\pm$ 2.34         | 43.42 $\pm$ 5.93        | 52.62 $\pm$ 1.78         |
| Bract width            | 16.21 $\pm$ 0.39         | 42.95 $\pm$ 1.46         | 31.2 $\pm$ 2.86         | 34.62 $\pm$ 0.93         |

strongly asymmetric postzygotic hybridization barrier has shaped the origin of the recently formed allopolyploid species *Mimulus peregrinus* and its triploid ancestor *M. xrobertsii* (= *M. guttatus*  $\times$  *M. luteus*). Thus, our study confirms early crossing experiments (Roberts, 1964; Parker, 1975) and adds to a growing list of exceptions to the “rule” of poorly performing paternal-excess crosses (Table 1; Appendix S1). Specifically, we showed that viable hybrids are considerably more likely to be produced when diploid *M. guttatus* is used as the maternal parent and tetraploid *M. luteus* as the paternal parent (MG  $\times$  ML), than in the opposite direction (ML  $\times$  MG). We showed that MG  $\times$  ML hybrids are larger than the parental taxa and the reciprocal hybrid ML  $\times$  MG. The MG  $\times$  ML hybrids also flowered earlier and produced more nectar than their *M. guttatus* parents. Motivated by this observed asymmetry of hybridization, we used previously available genomic sequences and developed novel genomic resources, including the first assembled chloroplast in the genus *Mimulus* and in the family Phrymaceae, to show that most individuals in natural populations of the triploid hybrid *M. xrobertsii* have *M. guttatus* as their maternal parent. As expected, *M. peregrinus*, which arose by genome duplication from *M. xrobertsii*, also has *M. guttatus* as its maternal parent. Beyond revealing the asymmetric origin of *M. xrobertsii* and *M. peregrinus*, our study shows the general potential of exploiting cytoplasmic genome data from both targeted and whole-genome sequence projects (Straub et al., 2012; Dodsworth, 2015) to investigate the evolution and speciation of polyploid and hybrid taxa.

**Pattern of hybridization asymmetry—Seed viability and germination**—Cytonuclear incompatibilities and differences in ploidy level, or a combination of both, can result in postzygotic asymmetric hybridization (Tiffin et al., 2001; Köhler et al., 2010; Scott et al., 2013). We currently cannot distinguish between these two classes of mechanisms to explain the observed pattern of asymmetric hybridization leading to *M. xrobertsii* and *M. peregrinus*. However, analysis of the morphology of hybrid seeds obtained in reciprocal crosses may offer some insight into the mechanisms of such asymmetry. For example, intraspecific hybridization between diploid and tetraploid *Arabidopsis thaliana* has shown that crosses with maternal excess (4 $\times$  to 2 $\times$ ) yield smaller seed sizes than the reciprocal paternal excess cross (2 $\times$  to 4 $\times$ ) (Scott et al., 2013). The seed size differences are thought to result from misregulation of imprinted genes that control growth and cell division in the endosperm (Köhler and Kradolfer, 2011; Schatlowski and Köhler, 2012; Lafon-Placette and Köhler, 2015). Paternal excess may cause endosperm over-proliferation and failed endosperm cellularization (resulting in large seeds),

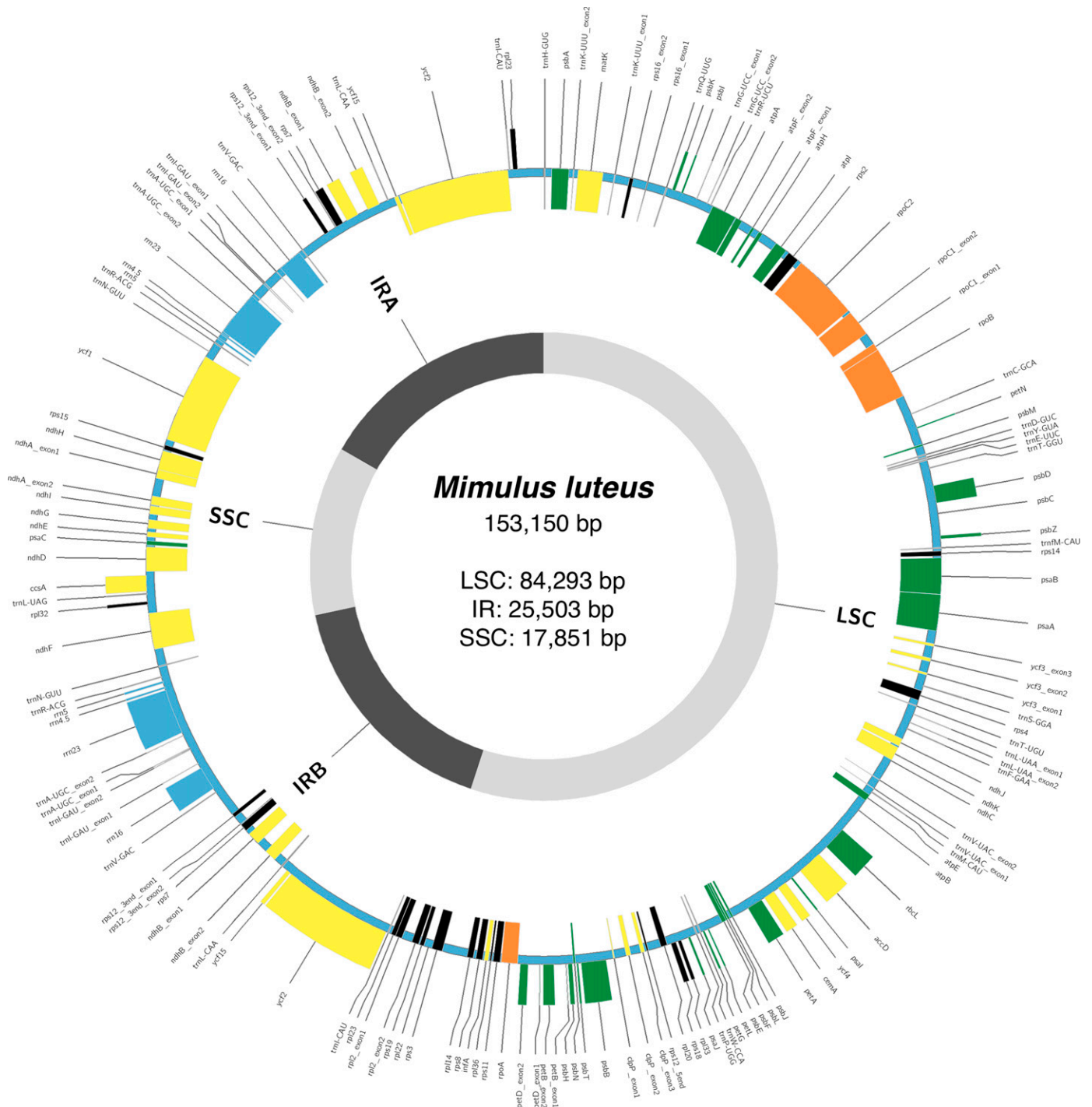
while maternal excess results in endosperm proliferation and small seeds (Haig and Westoby, 1991; Scott et al., 1998; Bushell et al., 2003; Scott et al., 2013).

As a first step toward understanding the mechanism of reproductive asymmetry in *Mimulus*, we compared seed size in the 16 controlled crosses used in the phenotypic experiment. These crosses included four accessions of each parental taxa and the reciprocal hybrid crosses (Appendix S2). We estimated seed size as seed area obtained from digital images of 300–600 seeds per cross type (online Appendix S10). Our results indicate that although both types of

hybrids have smaller seeds than the parental taxa produce, the ML  $\times$  MG cross produces larger (if mostly inviable) seeds than the reciprocal (and more viable) MG  $\times$  ML cross does (Appendix S10). This pattern suggests that interploidy crosses in *Mimulus* do not mimic the seed phenotypes of interploidy crosses in *Arabidopsis* (Scott et al., 2013). Further analyses of the morphology and histology of interploidy reciprocal crosses of *M. guttatus*  $\times$  *M. luteus* are needed to elucidate the mechanistic and genetic basis of postzygotic hybridization asymmetry in this group.

**Adult plant phenotype**—Most previous experimental studies comparing the viability of asymmetrically produced hybrids have focused on the immediate consequences of intrinsic postzygotic incompatibilities at the seed and seedling stage (seed number and seed germination). However, as we have shown, surviving hybrids may also differ in their adult phenotypes depending on the cross direction. Although our sample size is small, our results show that paternal-excess *Mimulus* triploids possessed characteristics potentially associated with high extrinsic fitness, including larger plant size. Paternal excess hybrids also flowered earlier and produced more nectar than maternal-excess triploids, but the direct effects of these traits on hybrid fitness are expected to be negligible given that *M. xrobertsii* is sexually sterile (Roberts, 1964; Stace, 2010), and this sterility does not differ in synthetic hybrids created using reciprocal crosses (Roberts, 1964, p. 73). A major gap in our knowledge is whether phenotypic differences among reciprocal hybrids observed in greenhouse conditions translate to performance differences in the field, and further studies on the ecology and fitness of reciprocal hybrids and their parents are needed. It would also be important to compare the morphology of triploid *M. xrobertsii* and its allohexaploid derivative *M. peregrinus* to assess the effects of genome duplication on the phenotype. Preliminary field observations suggest that allopolyploids are larger in size for some morphological traits such as flower size (M. Vallejo-Marín, personal observations). The combination of *Mimulus* genomic and transgenic resources (Wu et al., 2008; Hellsten et al., 2013; Yuan et al., 2013; Twyford et al., 2015) with a rich ecological knowledge base, makes the naturalized hybrids and neopolyploids of the British Isles an exciting model for connecting molecular mechanisms of interploidy asymmetry to observed patterns of speciation in the wild.

**Genetic ancestry of *M. xrobertsii* and *M. peregrinus***—Regardless of mechanism, our study seems to be an exception to the general pattern in which viable interploidy hybrids are more likely to be produced when the maternal parent is of higher ploidy level than the

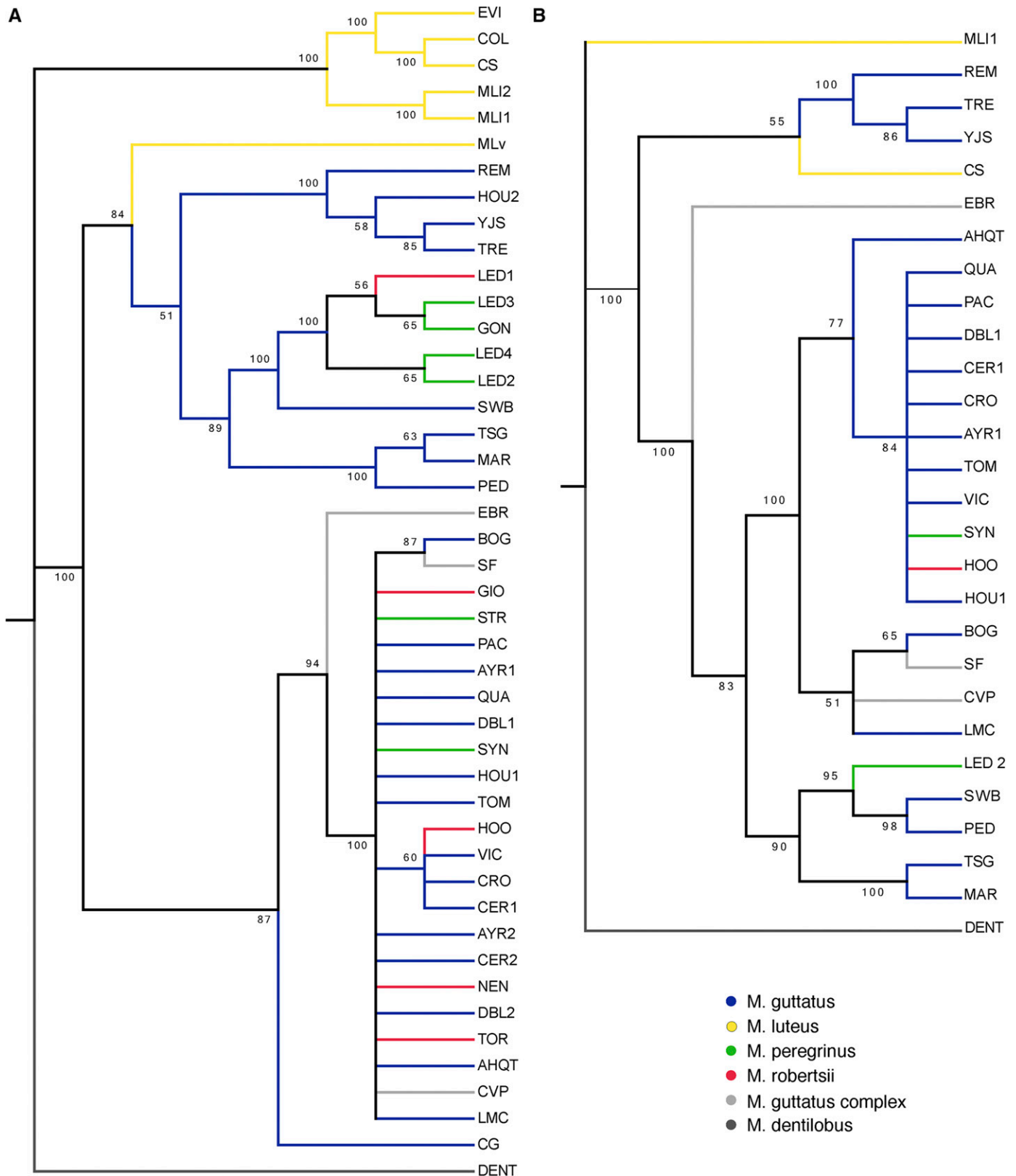


**FIGURE 3** Circular plot of *Mimulus luteus* var. *luteus* chloroplast genome. The plastome is 153,150 bp with a 84,293-bp large single copy (LSC) region, a 17,851-bp small single copy (SSC) region, and two 25,503-bp inverted repeat regions (IRB and IRA). Photosystem genes are in green, RNA polymerase subunits are in orange, ribosomal genes are in black, other protein encoding genes are in yellow, rRNAs are in blue, and tRNAs are in black.

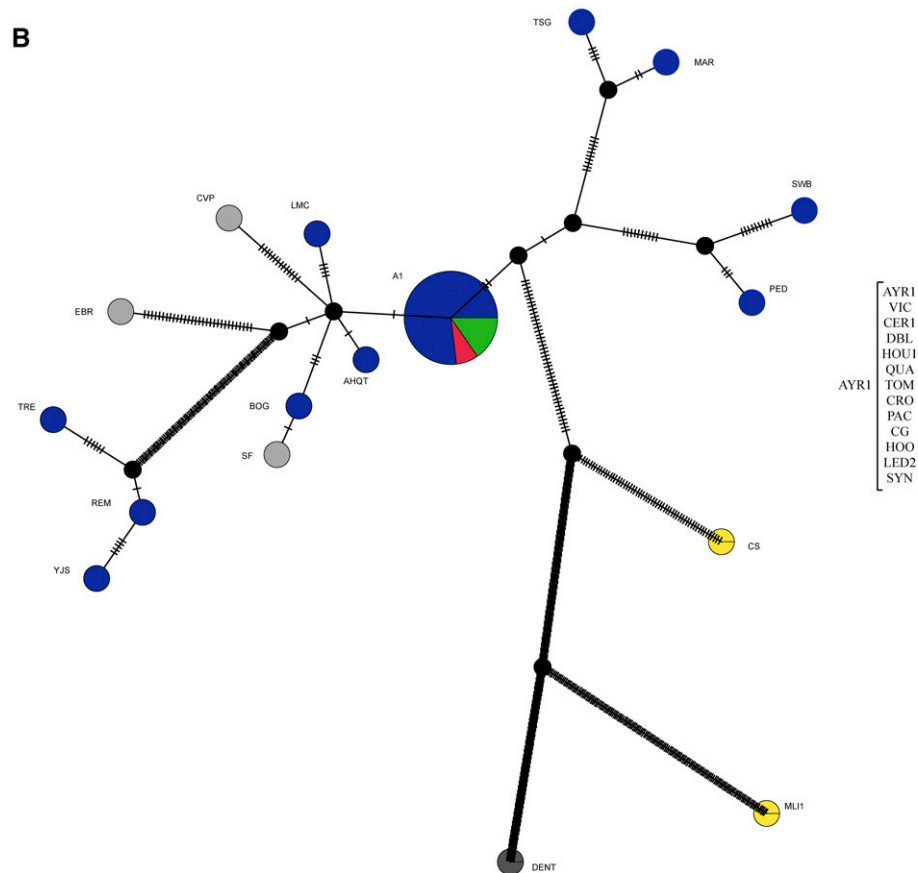
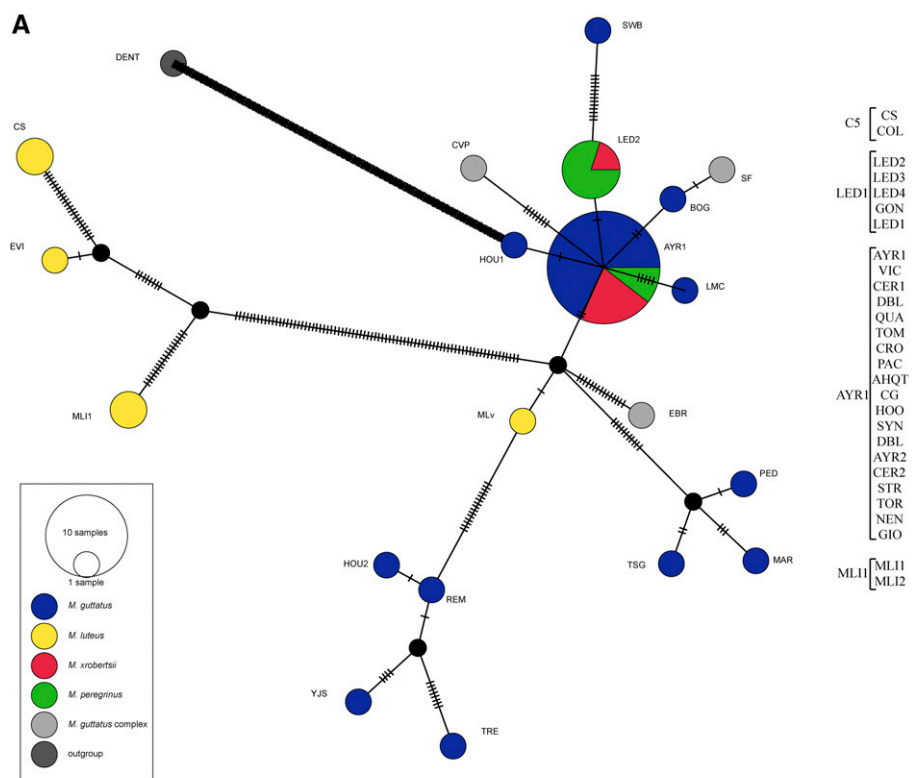
paternal one (maternal excess; Stebbins, 1957; Ramsey and Schemske, 1998). Both chloroplast and mitochondrial genomic analyses clearly show that the triploid and hexaploid individuals cluster more closely with *M. guttatus* than with any native or introduced *M. luteus* (Figs. 4–6). The genomic data also show that even whole-genome sequences are unable to completely separate *M. guttatus* and *M. luteus*, which may reflect incomplete lineage sorting between

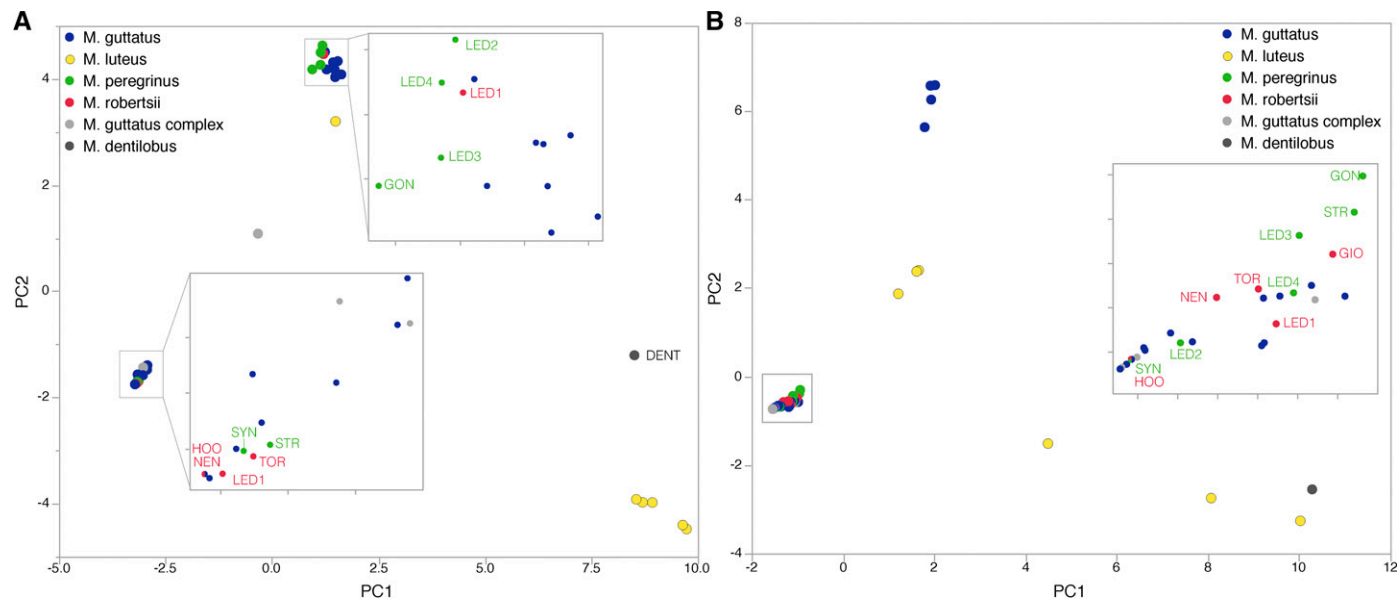
these relatively closely related taxa (Maddison and Knowles, 2006). Moreover, the origin of *M. luteus* itself is unknown, but nuclear genomic data suggest that this taxon has a hybrid origin (Vallejo-Marín et al., 2015), perhaps with a *M. guttatus*-like ancestor (Mukherjee and Vickery, 1962). If the putative allopolyploid *M. luteus* has had multiple origins, then the simple expectation of reciprocal monophyly for the plastid genomes of *M. luteus* and





**FIGURE 4** Maximum likelihood tree of *Mimulus* species and populations based on alignments of (A) 86,843 bp from chloroplast genomes and (B) 402,717 bp from mitochondrial genomes. Bootstrap values are indicated. Nodes with less than 50% bootstrap support were collapsed. Samples genotyped using sequence-capture ( $n = 16$ ) are excluded from the mitochondrial analysis due to low genotyping success. Trees are rooted with *M. dentilobus*. Color-coding is based on species identity.





**FIGURE 6** Principal component analysis of (A) 929 SNPs from the chloroplast and (B) 1454 SNPs from the mitochondria. Only sites successfully genotyped in 41/45 samples were included in this analysis. The chloroplast data set excludes one individual (CG) with low genotyping success for the analyzed SNPs. The mitochondrial data set excludes 16 samples obtained through sequence capture due to low genotyping success (see Results). A fully annotated version of this figure is available in online Appendices S8 and S9.

*M. guttatus* may not be met. Solving the mystery of the origin of *M. luteus* and its relationship with *M. guttatus* and related taxa will require detailed analyses of natural populations throughout the native South American range.

Our survey of hybrid *Mimulus* populations across the British Isles using three mitochondrial loci also supports *M. guttatus* as the maternal parent of *M. ×robertsii* and *M. peregrinus*. Of 114 triploid *M. ×robertsii* from 13 populations and 27 individuals from three populations of the hexaploid *M. peregrinus* that were genotyped at these three mitochondrial loci, all but three individuals of *M. ×robertsii* from a single population had maternal haplotypes that could be confidently assigned to *M. guttatus*. The three individuals of *M. ×robertsii* (all from population NEN) that had *M. luteus* mitochondrial haplotypes (Table 3) pose an exception to the general pattern detected here. An *M. luteus*-like maternal haplotype in hybrids could be explained in a number of ways. First, it could simply be due to contamination during genotyping. However, we re-extracted and re-genotyped these samples and confirmed the mitochondrial haplotype. Second, the three samples could have been mistakenly identified as *M. ×robertsii* but instead belong to *M. luteus*. We think this possibility is also unlikely as plants were morphologically identified on site, and we also conducted surveys of this population (NEN) in 2010, 2011, and 2013, and *M. luteus* has never been observed in this or nearby populations. Third, the haplotypes in these three hybrids could arise via introgression from *M. luteus*. Introgression across ploidy barriers

has been shown for nuclear markers in other species (Chapman and Abbott, 2010) and could possibly occur at the mitochondrial level as well. Although we cannot rule out this possibility, the lack of local *M. luteus* (the nearest known *M. luteus* natural population is ~100 km north) and the principal fact that hybrids are not known to set seed (i.e., are almost completely sterile; Roberts, 1964; Vallejo-Marín, 2012) makes the introgression scenario unlikely. Fourth, the “*M. luteus*” haplotype in hybrids could be explained by incomplete lineage sorting of ancestral polymorphisms between *M. guttatus* and *M. luteus* (Twyford and Ennos, 2012). We cannot currently rule out ancestral polymorphism, which remains a distinct possibility given the observations that a single individual of *M. guttatus* (DBL) bears a mitochondrial haplotype otherwise characteristic of *M. luteus* (Table 2). Similarly, the mitochondrial haplotype for the British individual of *M. luteus* s.l. (CS) nested within the *M. guttatus* clade also suggests incomplete lineage sorting. Given the limited sampling of parental genotypes, incomplete lineage sorting remains a very likely possibility. Fifth, homoplasy resulting from the sequences converging on a similar haplotype (especially if the sequences are not that distinct) could also explain this pattern. Finally, a simpler explanation may be that although the majority of hybrids are produced when *M. guttatus* is the maternal parent, occasionally, the opposite cross can yield viable offspring. Indeed, our experimental results show that viable hybrids in the ML × MG direction are produced albeit with low probability (Fig. 1).

**FIGURE 5** Matrilineal haplotype network (statistical parsimony; TCS) of (A) chloroplast and (B) mitochondrial SNPs. The chloroplast analysis is based on 493 sites and the mitochondrial on 784 sites. Sites with more than 5% missing data were excluded. Hash marks on lines are mutational steps. The haplotype circles are labeled with one individual, and all other individuals with that haplotype are listed to the right. The mitochondrial data set excludes 16 samples obtained through sequence capture due to low genotyping success (see Results). Names for individual samples are only shown here for *M. ×robertsii* and *M. peregrinus*.



**Skimming of cytoplasmic genomes**—Acquisition of complete organellar genomes is increasingly feasible with the advent of “genome skimming” techniques associated with next-generation sequencing projects. In fact, mitochondrial sequences usually represent 1–5% of the reads obtained in such projects (Steele et al., 2012; Smith, 2015) and are often discarded as contamination during nuclear genomic analyses. Our study showed that data generated by whole-genome sequencing contain enough cytoplasmic “contamination” to yield between 1515- and 208-fold mean coverage of the chloroplast and mitochondria, respectively. Even the approach of sequence capture, in which specific (nuclear) regions were targeted (Vallejo-Marín et al., 2015) contained sufficient reads to build large contigs of both chloroplast and mitochondrial genes (86,843 bp and 402,717 bp, respectively). As expected, however, the sequence capture data did not allow us to genotype all the same sites in all individuals. The limited coverage of specific sites was particularly acute for the mitochondrial genome, which reduced its utility in some of the analyses here. Our results suggest that using off-targeted reads from sequence capture approaches may be sufficient to generate moderate numbers of genetic markers (e.g., Fig. 6), but this approach is of limited utility when coverage of large numbers of sites across multiple individuals is required (e.g., Figs. 4, 5).

**The *Mimulus* chloroplast genome**—Shallow whole-genome sequencing (also referred to as genome skimming) allowed us to reconstruct the chloroplast genome of *M. luteus*. The structure and gene content of this genome follows what has been seen for other taxa in Lamiales (Nazareno et al., 2015), including that of *Utricularia gibba* (Ibarra-Laclette et al., 2013), a species known for a diminutive nuclear genome. The relatively constant architecture and gene content identified across multiple lineages of the order Lamiales suggests that there may be selection for maintenance of gene content and structure relative to other eudicots lineages like saguaro (Sanderson et al., 2015) and legumes (Saski et al., 2005). A more comprehensive taxon sampling of chloroplast genomes is needed throughout the Lamiales and all angiosperm to truly understand the dynamics and evolution of this vital organelle.

The use of genome skimming has repeatedly proven to be a useful technique for acquiring either complete gene sets (e.g., Washburn et al., 2015) or full plastomes (e.g., Bock et al., 2013). Here we demonstrated the utility of a single plastome and mitochondrial genome as anchors to identify organellar genome sequences rapidly from genome skimming data. Though full plastid and mitochondrial genome sequences were not recovered, a large proportion of each genome was, and those regions were identified to annotated features and alignable, providing resolution of relationships at the population level. Future work in the group will include de novo assembly of whole chloroplast genomes across the genus providing valuable resources for population-level studies of hybridization and polyploidy.

## CONCLUSIONS

The value of genome skimming for general phylogenetic analyses is well recognized (e.g., Hahn et al., 2013). We demonstrate that the analysis of the entire mitochondrial and chloroplast genomes can yield information to infer the maternal ancestry of a hybrid and neo-allopolyploid derived from recently diverged taxa (Beardsley

and Olmstead, 2002; Nie et al., 2006). However, our study also shows that even whole-genome sequences might only imperfectly resolve phylogenetic relationships between closely related species. Therefore, genomic analysis of hybrid ancestry can be significantly strengthened with data from experimental crosses as done here. Exploration of reproductive isolation asymmetry across the plant kingdom through genome-skimming may allow for a better understanding of the origin of hybrid and allopolyploid species. Connecting the genomic and molecular underpinnings of reproductive asymmetry to the morphological and ecological consequences of asymmetric barriers will ultimately generate a better solution to the puzzle of the evolutionary success of allopolyploids.

## ACKNOWLEDGEMENTS

The authors thank H. E. O'Brien for comments and suggestions on haplotype analysis. The authors thank two anonymous reviewers for comments on a previous version of this manuscript. This work was supported in part by a UK Natural Environment Research Council grant (NERC, NE/J012645/1), a NERC Biomolecular Analysis Facilities grant (NBAF710) and a Carnegie Trust Travel grant to M.V.M. and a Suzann Wilson Matthews Research Award to J.R.P.

## LITERATURE CITED

- Auwerwa, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, et al. 2013. From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics* 11: 11.10.11–11.10.33.
- Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.
- Barker, M. S., N. Arrigo, A. E. Baniaga, Z. Li, and D. A. Levin. 2016. On the relative abundance of autopolyploids and allopolyploids. *New Phytologist* 210: 391–398.
- Beardsley, P. M., and R. G. Olmstead. 2002. Redefining Phrymaceae: The placement of *Mimulus*, tribe Mimuleae and Phryma. *American Journal of Botany* 89: 1093–1102.
- Behrend, A., A. Gluschak, A. Przybyla, and A. Hohe. 2015. Interploid crosses in heather (*Calluna vulgaris*). *Scientia Horticulturae* 181: 162–167.
- Bock, D. G., N. C. Kane, D. P. Ebert, and L. H. Rieseberg. 2013. Genome skimming reveals the origin of the Jerusalem artichoke tuber crop species: Neither from Jerusalem nor an artichoke. *New Phytologist* 201: 1021–1030.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: btu170.
- Burton, T. L., and B. C. Husband. 2000. Fitness differences among diploids, tetraploids, and their triploid progeny in *Chamerion angustifolium*: Mechanisms of inviability and implications for polyploid evolution. *Evolution* 54: 1182–1191.
- Bushell, C., M. Spielman, and R. J. Scott. 2003. The basis of natural and artificial postzygotic hybridization barriers in *Arabidopsis* species. *Plant Cell* 15: 1430–1442.
- Castro, S., Z. Münzbergová, J. Raabová, and J. Loureiro. 2011. Breeding barriers at a diploid–hexaploid contact zone in *Aster amellus*. *Evolutionary Ecology* 25: 795–814.
- Chapman, M. A., and R. J. Abbott. 2010. Introgression of fitness genes across a ploidy barrier. *New Phytologist* 186: 63–71.
- Cheung, K. W., F. M. Razeq, C. A. Sauder, T. James, and S. L. Martin. 2015. Bidirectional but asymmetrical sexual hybridization between *Brassica carinata* and *Sinapis arvensis* (Brassicaceae). *Journal of Plant Research* 128: 469–480.
- Comai, L. 2005. The advantages and disadvantages of being polyploid. *Nature Reviews. Genetics* 6: 836–846.

- Cooley, A. M., and J. H. Willis. 2009. Genetic divergence causes parallel evolution of flower color in Chilean *Mimulus*. *New Phytologist* 183: 729–739.
- DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43: 491–498.
- Dinu, I., R. Hayes, R. Kynast, R. Phillips, and C. Thill. 2005. Novel inter-series hybrids in *Solanum*, section *Petota*. *Theoretical and Applied Genetics* 110: 403–415.
- Dodsworth, S. 2015. Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science* 20: 525–527.
- Doyle, J. J., and J. L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus (San Francisco, Calif.)* 12: 13–15.
- Field, D., D. Ayre, R. Whelan, and A. Young. 2011. Patterns of hybridization and asymmetrical gene flow in hybrid zones of the rare *Eucalyptus aggregata* and common *E. rubida*. *Heredity* 106: 841–853.
- Grant, A. L. 1924. A monograph of the genus *Mimulus*. *Annals of the Missouri Botanical Garden* 11: 99–389.
- Greiner, R., and C. Oberprieler. 2012. The role of inter-ploidy block for reproductive isolation of the diploid *Leucanthemum pluriflorum* Pau (Compositae, Anthemideae) and its tetra- and hexaploid relatives. *Flora—Morphology, Distribution, Functional Ecology of Plants* 207: 629–635.
- Hahn, C., L. Bachmann, and B. Chevreaux. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—A baiting and iterative mapping approach. *Nucleic Acids Research* 41: e129.
- Haig, D., and M. Westoby. 1989. Parent-specific gene expression and the triploid endosperm. *American Naturalist* 134: 147–155.
- Haig, D., and M. Westoby. 1991. Genomic imprinting in endosperm: Its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 333: 1–13.
- Hellsten, U., K. Wright, J. Jenkins, S. Shu, Y. Yuan, S. Wessler, J. Schmutz, et al. 2013. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proceedings of the National Academy of Sciences, USA* 110: 19478–19482.
- Hersch-Green, E. I. 2012. Polyploidy in Indian paintbrush (*Castilleja*; *Orobanchaceae*) species shapes but does not prevent gene flow across species boundaries. *American Journal of Botany* 99: 1680–1690.
- Holleley, C. E., and P. G. Geerts. 2009. Multiplex Manager 1.0: A cross-platform computer program that plans and optimizes multiplex PCR. *BioTechniques* 46: 511–517.
- Hothorn, T., F. Bretz, and P. Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal. Biometrische Zeitschrift* 50: 346–363.
- Husband, B. C. 2004. The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biological Journal of the Linnean Society* 82: 537–546.
- Ibarra-Laclette, E., E. Lyons, G. Hernández-Guzmán, et al. 2013. Architecture and evolution of a minute plant genome. *Nature* 498: 94–98.
- Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* 27: 3070–3071.
- Katoh, K., K.-i. Kuma, H. Toh, and T. Miyata. 2005. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* 33: 511–518.
- Kelly, A. J., and J. H. Willis. 1998. Polymorphic microsatellite loci in *Mimulus guttatus* and related species. *Molecular Ecology* 7: 769–774.
- Köhler, C., and D. Kradolfer. 2011. Epigenetic mechanisms in the endosperm and their consequences for the evolution of flowering plants. *Biochimica et Biophysica Acta. Gene Regulatory Mechanisms* 1809: 438–443.
- Köhler, C., O. M. Scheid, and A. Erilova. 2010. The impact of the triploid block on the origin and evolution of polyploid plants. *Trends in Genetics* 26: 142–148.
- Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra. 2009. Circos: An information aesthetic for comparative genomics. *Genome Research* 19: 1639–1645.
- Lafon-Placette, C., and C. Köhler. 2015. Epigenetic mechanisms of postzygotic reproductive isolation in plants. *Current Opinion in Plant Biology* 23: 39–44.
- Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.
- Levin, D. A. 2002. The role of chromosomal change in plant evolution. Oxford University Press on Demand, Oxford, UK.
- Maddison, W. P., and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55: 21–30.
- Mallet, J. 2007. Hybrid speciation. *Nature* 446: 279–283.
- Marçais, G., and C. Kingsford. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27: 764–770.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernysky, K. Garimella, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297–1303.
- Miyashita, T., and Y. Hoshino. 2015. Interploid and intraploid hybridizations to produce polyploid Haskap (*Lonicera caerulea* var. *emphyllocalyx*) plants. *Euphytica* 201: 15–27.
- Mower, J. P., A. L. Case, E. R. Floro, and J. H. Willis. 2012. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. *Genome Biology and Evolution* 4: 670–686.
- Mukherjee, B. B., and R. K. Vickery. 1962. Chromosome counts in the section *Simiolus* of the genus *Mimulus* (Scrophulariaceae). V. The chromosomal homologies of *M. guttatus* and its allied species and varieties. *Madroño* 16: 141–172.
- Nazareno, A. G., M. Carlsen, and L. G. Lohmann. 2015. Complete chloroplast genome of *Tanaecium tetragonolobum*: The first Bignoniaceae plastome. *PLoS One* 10: e0129930.
- Nghiem, Q., J. Harbard, C. Harwood, A. Griffin, T. Ha, and A. Koutoulis. 2013. Pollen–pistil interactions between autotetraploid and diploid *Acacia mangium* and diploid *A. auriculiformis*. *Journal of Tropical Forest Science* 25: 96–110.
- Nie, Z., H. Sun, P. Beardsley, R. Olmstead, and J. Wen. 2006. Evolution of biogeographic disjunction between eastern Asia and eastern North America in *Phryma* (Phrymaceae). *American Journal of Botany* 93: 1343–1356.
- Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401–437.
- Parker, P. F. 1975. *Mimulus* in Great Britain—A cytotaxonomic note. *New Phytologist* 74: 155–160.
- Petit, R. J., C. Bodénès, A. Ducouso, G. Roussel, and A. Kremer. 2004. Hybridization as a mechanism of invasion in oaks. *New Phytologist* 161: 151–164.
- Preston, C. D., D. A. Pearman, and T. D. Dines. 2002. New atlas of the British and Irish flora. An atlas of the vascular plants of Britain, Ireland, the Isle of Man and the Channel Islands. Oxford University Press, Oxford, UK.
- Puzey, J., and M. Vallejo-Marín. 2014. Genomics of invasion: Diversity and selection in introduced populations of monkeyflowers (*Mimulus guttatus*). *Molecular Ecology* 23: 4472–4485.
- Rhymer, J. M., and D. Simberloff. 1996. Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* 27: 83–109.
- Qian, J., J. Song, H. Gao, Y. Zhu, J. Xu, X. Pang, H. Yao, et al. 2013. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One* 8: e57607.
- R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website <http://www.R-project.org/>.
- Ramsey, J., and D. W. Schemske. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* 29: 467–501.
- Roberts, R. 1964. *Mimulus* hybrids in Britain. *Watsonia* 6: 70–75.
- Roccaforte, K., S. E. Russo, and D. Pilson. 2015. Hybridization and reproductive isolation between diploid *Erythronium mesochoreum* and its tetraploid congener *E. albidum* (Liliaceae). *Evolution* 69: 1375–1389.

- Sabara, H. A., P. Kron, and B. C. Husband. 2013. Cytotype coexistence leads to triploid hybrid production in a diploid–tetraploid contact zone of *Chamerion angustifolium* (Onagraceae). *American Journal of Botany* 100: 962–970.
- Sanderson, M. J., D. Copetti, A. Búrquez, et al. 2015. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted repeat. *American Journal of Botany* 102: 1115–11127.
- Saski, C., S.-B. Lee, H. Daniell, T. C. Wood, J. Tomkins, H.-G. Kim, and R. K. Jansen. 2005. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Molecular Biology* 59: 309–322.
- Schatlowski, N., and C. Köhler. 2012. Tearing down barriers: Understanding the molecular mechanisms of interploidy hybridizations. *Journal of Experimental Botany* 63: 6059–6067.
- Scott, R. J., M. Spielman, J. Bailey, and H. G. Dickinson. 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* 125: 3329–3341.
- Scott, R. J., J. L. Tratt, and A. Bolbol. 2013. Seed development in interploidy hybrids. In Z. J. Chen and J. A. Birchler [eds.], *Polyploid and hybrid genomics*, 271–290. John Wiley, Oxford, UK.
- Sekine, D., T. Ohnishi, H. Furuumi, A. Ono, T. Yamada, N. Kurata, and T. Kinoshita. 2013. Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. *Plant Journal* 76: 792–799.
- Silverside, A. J. 1990. A new binomial in *Mimulus*. *Watsonia* 18: 210–212.
- Silverside, A. J. 1998. *Mimulus* section *Simiolus*. In T. Rich and A. Jermy [eds.], *Plant crib*, 259–261. Botanical Society of the British Isles, London, UK.
- Smith, D. R. 2015. The past, present and future of mitochondrial genomics: Have we sequenced enough mtDNAs? *Briefings in Functional Genomics*.
- Soltis, D. E., M. C. Segovia-Salcedo, I. Jordon-Thaden, L. Majure, N. M. Miles, E. V. Mavrodiev, W. Mei, et al. 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytologist* 202: 1105–1117.
- Stace, C. A. 2010. *New flora of the British Isles*. Cambridge University Press, Cambridge, UK.
- Stace, C. A., C. D. Preston, and D. A. Pearman. 2015. *Hybrid flora of the British Isles*. BSBI Publications, Bristol, UK.
- Stamatakis, A. 2006. RAXML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
- Stebbins, G. L. 1957. Self fertilization and population variability in higher plants. *American Naturalist* 91: 337–354.
- Steele, P. R., K. L. Hertweck, D. Mayfield, M. R. McKain, J. Leebens-Mack, and J. C. Pires. 2012. Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae. *American Journal of Botany* 99: 330–348.
- Stoute, A. I., V. Varenko, G. J. King, R. J. Scott, and S. Kurup. 2012. Parental genome imbalance in *Brassica oleracea* causes asymmetric triploid block. *Plant Journal* 71: 503–516.
- Straub, S. C., M. Parks, K. Weitemier, M. Fishbein, R. C. Cronn, and A. Liston. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.
- Tiffin, P., S. Olson, and L. C. Moyle. 2001. Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society of London, B, Biological Sciences* 268: 861–867.
- Toonen, R. J., and S. Hughes. 2001. Increased throughput for fragment analysis on an ABI Prism® 377 automated sequencer using a membrane comb and STRand software. *BioTechniques* 31: 1320–1325.
- Twyford, A., and R. Ennos. 2012. Next-generation hybridization and introgression. *Heredity* 108: 179–189.
- Twyford, A. D., M. A. Streisfeld, D. B. Lowry, and J. Friedman. 2015. Genomic studies on the nature of species: Adaptation and speciation in *Mimulus*. *Molecular Ecology* 24: 2601–2609.
- Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. 2012. Primer3—New capabilities and interfaces. *Nucleic Acids Research* 40: e115.
- Vallejo-Marín, M. 2012. *Mimulus peregrinus* (Phrymaceae): A new British allopolyloid species. *PhytoKeys* 14: 1–14.
- Vallejo-Marín, M., R. J. Buggs, A. M. Cooley, and J. R. Puzey. 2015. Speciation by genome duplication: Repeated origins and genomic composition of the recently formed allopolyloid species *Mimulus peregrinus*. *Evolution* 69: 1487–1500.
- Vallejo-Marín, M., and G. C. Lye. 2013. Hybridisation and genetic diversity in introduced *Mimulus* (Phrymaceae). *Heredity* 110: 111–122.
- Van Laere, K., D. Hermans, L. Leus, and J. Van Huylenbroeck. 2015. Interspecific hybridisation within *Buxus* spp. *Scientia Horticulturae* 185: 139–144.
- Von Bohlen, C. 1995. El género *Mimulus* L. (Scrophulariaceae) en Chile. *Gayana Botanica* 52: 7–28.
- Washburn, J. D., J. C. Schnable, G. Davidse, and J. C. Pires. 2015. Phylogeny and photosynthesis of the grass tribe Paniceae. *American Journal of Botany* 102: 1493–1505.
- Watkins, A. 1932. Hybrid sterility and incompatibility. *Journal of Genetics* 25: 125–162.
- Watson, J. M. 1989. 'Andean Nymph,' a pink *Mimulus*. Thompson & Morgan, Ipswich, UK.
- Weitemier, K., S. C. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2.9: apps.1400042.
- Wu, C. A., D. B. Lowry, A. M. Cooley, K. M. Wright, Y.-W. Lee, and J. H. Willis. 2008. *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100: 220–230.
- Wyman, S. K., R. K. Jansen, and J. L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.
- Yuan, Y.-W., J. M. Sagawa, R. C. Young, B. J. Christensen, and H. D. Bradshaw. 2013. Genetic dissection of a major anthocyanin QTL contributing to pollinator-mediated reproductive isolation between sister species of *Mimulus*. *Genetics* 194: 255–263.
- Zhang, H., C. Li, H. Miao, and S. Xiong. 2013. Insights from the complete chloroplast genome in the evolution of *Sesamum indicum* L. *PLoS One* 8: e80508.